Journal of
Experimental Psychology

VOL. 32, NO. 3 MARCH, 1943

RETROACTIVE INHIBITION AS A FUNCTION OF DEGREE OF INTERPOLATED LEARNING

BY LELAND E. THUNE AND BENTON J. UNDERWOOD

State University of Iowa[1]

I. INTRODUCTION

At least three different sets of operations have been applied to define the extent of interpolated activity as it influences retroactive inhibition (RI). The interpolated activity can be varied (a) as by McGeoch (**4**), in terms of the length of the list which is presented, (b) as by Twining (**12**), in terms of the number of lists of constant length, and (c) as by McGeoch (**3**), Melton and Irwin (**8**) and others, in terms of the number of repetitions of a single list. As might be expected, different results have been obtained in the various studies, depending upon which of the above operations has been used. The present study is concerned with RI as determined by the extent of interpolation in the last sense, *i.e.*, the number of repetitions of a single list.

McGeoch (**3**) has reported the first systematic study on RI as a function of the number of repetitions of the interpolated material. Eight nonsense syllables made up the original and interpolated lists, with the first list always presented for 11 trials. The interpolated lists were presented for 6, 11, 16, 21, or 26 repetitions. Results showed that RI increased initially with additions in the number of trials on the interpolated list, but leveled off as complete mastery was attained. Further repetitions had little influence. McGeoch's data, however, showed a tendency for RI to decrease slightly with high degrees of learning on the interpolated list, and other work by the same experimenter (**3**) demonstrated this same tendency.

Melton and Irwin (**8**) used 5 trials for the original learning on lists of 18 nonsense syllables, and 5, 10, 20, or 40 trials for the interpolated learning. Recall and relearning after 30 min. showed that retroactive

[1] We wish to thank Prof. Kenneth W. Spence and Dr. Claude E. Buxton for their valuable suggestions in the preparation of this manuscript.

effects were greatest for the condition involving a 10-trial interpolation and least for 5 and 40 trials. This tendency for RI to decrease with 40 interpolated trials was more apparent in the number of trials required to relearn the list to mastery than it was in the recall scores.

These studies, together with Lahey's work (2) in the classroom, show unequivocally that RI increases with increases in the degree of interpolation, maximal RI obtaining with intermediate degrees of interpolated learning. Continued repetition produces a leveling of the inhibition curve with decreases at extreme overlearning.

To date, no studies on this function in which paired associates learning was involved have been reported. The present study was designed to test the generality of these previous findings through the use of paired two-syllable adjectives, with 2, 5, 10, and 20 repetitions making up the degrees of interpolation. In addition, this experiment would provide a check on the Melton-Irwin theoretical curve (8) which has been the basis for the postulation of a two-factor theory of RI.

II. Method

Subjects and Materials.—The 24 volunteer *Ss* who participated in this experiment were all male undergraduates. These *Ss* learned lists of 10 paired associates by the anticipation method. The original and interpolated lists were of the A–B, A–C relationship, the same stimulus word appearing in both the original and interpolated lists. Thus, two different response items had to be associated with the same stimulus term.

All adjectives contained two syllables and were made into lists according to the following rules:

1. The initial letter of any adjective in a given list was never used more than once in that list.
2. A given suffix might appear more than once in a list, but each of the 4 adjectives of 2 successive pairs had a different suffix.
3. The initial letters of the items of a given pair taken in the order of the items (A–B) were not successive letters of the alphabet.
4. The adjectives in a given pair were neither synonyms nor antonyms.
5. Even though the two adjectives of a given pair were neither synonyms nor antonyms, they were not paired if they obviously suggested a paired usage in common parlance, *e.g.*, 'perfect-lovely.'
6. No two adjectives in a pair were sonorous, *e.g.*, 'lawless-awful.'
7. In the construction of the interpolated lists, the above rules obtained.

In order to enforce strict paired associates learning and to minimize serial associations between pairs, 5 different random orders of presentation were used. Hence in every group of 5 trials the orders were repeated.

The lists of paired words, typed in the 5 orders on a single tape, were presented on a modified Hull-type memory drum. The drum was adjusted to provide a 2-sec. anticipation period. An 8-sec. interval, during which the drum kept moving but only blank spaces appeared, followed each trial.

The instructions to the *Ss* did not disclose the true nature and purpose of the experiment, nor were the *Ss* informed ahead of time as to what the experimental condition on any given day would be.

Specific Conditions.—The 4 degrees of interpolated learning in this experiment were 2, 5, 10, and 20 repetitions of the list. The sequence of conditions used to measure the influence of these degrees of interpolation on RI is given in Table I.

Original learning always consisted of 5 trials, with recall and relearning after 20 min. With 2, 5, 10, and 20 repetitions on the interpolated lists, the rest periods were diminished to 17½,

15, 11, and 3 min., respectively. All of these major rest periods were filled by having *Ss* rate a series of cartoons. Relearning was always carried to 2 successive errorless trials.

Condition F was included in the design, not for the data which it would yield, but for certain methodological reasons. The prime purpose for this condition is that an experiment with 6 conditions can be counterbalanced more effectively than one with 5 conditions. In addition, the general practice level of the *Ss* was raised 'one day.'

It is necessary to keep the practice level of all *Ss* relatively constant on any given day, regardless of the prior conditions under which they may have served. To accomplish this, 25 repetitions were given on each experimental day, exclusive of the trials to learn and relearn the original list. Thus, the opportunity for practice was constant for each *S* for each day. The additional trials necessary to maintain this level of practice always came after relearning the original list. On condition A (no interpolation) the 25 trials were given after relearning the original list, and on conditions B, C, D, and E, 23, 20, 15, and 5 repetitions, respectively. As an approximation, 15 trials were given on condition F.

Before entering upon the experiment proper, each *S* was given 2 practice periods under conditions simulating those of the actual experiment. During these practice sessions *E* interrupted the learning at any time to point out failures to follow instructions.

TABLE I

SEQUENCE OF EXPERIMENTAL CONDITIONS USED TO MEASURE RI AFTER 2, 5, 10, AND 20 INTERPOLATED TRIALS

Condition	Original Learning	Rest	Interpolated Learning	Rest	Relearning Original List
A	5 trials	—	20 min. rest	—	to 2 perfect
B	5 trials	30″	2 trials	17½′	to 2 perfect
C	5 trials	30″	5 trials	15′	to 2 perfect
D	5 trials	30″	10 trials	11′	to 2 perfect
E	5 trials	30″	20 trials	3′	to 2 perfect
F	2 perfect	30″	20 min. rest	—	to 2 perfect

Each *S* served in all 6 experimental conditions. On each successive experimental day the same lists were learned by every *S* regardless of the conditions occurring on that day. The experimental conditions were counterbalanced to equalize as far as possible the practice effects which accrue with continued learning, and to distribute as evenly as possible over all conditions any unknown differential effects of the lists used. Six sets of experimental lists were employed.

The counterbalancing system used adhered to the following rules: (a) each *S* learned under a different order of conditions; (b) each condition appeared in each serial position 4 times; (c) each condition preceded and followed every other condition a total of 4 times.

III. RESULTS

Degree of Original Learning.—It is necessary to show, before making inter-condition comparisons, that the 5 repetitions used for the original learning resulted in comparable levels of acquisition for all conditions. The data required for such comparisons are given in Table II. When degree of learning is measured in terms of number of correct anticipations given during all the 5 trials, the largest difference is between conditions C and D. This difference, 1.42 ± 1.85, is only .77 times its sigma. When degree of original learning is measured in terms of the number of correct anticipations occurring on the fifth trial (last original learning trial), the largest difference obtains between conditions E and F, being .38 ± .58 (t = .66). Thus, the degree of

TABLE II

MEAN NUMBER OF CORRECT ANTICIPATIONS DURING ALL 5 ORIGINAL LEARNING TRIALS AND ON THE FIFTH ORIGINAL LEARNING TRIAL ALONE

Condition	All 5 OL Trials		5th Trial Alone	
	M	σM	M	σM
A (0 IL trials)	14.00	1.84	4.96	.54
B (2 IL trials)	13.92	1.55	5.08	.44
C (5 IL trials)	13.71	1.65	4.88	.47
D (10 IL trials)	15.13	1.83	5.08	.58
E (20 IL trials)	14.54	2.00	4.83	.58
F (0 IL trials)	14.13	1.51	5.21	.51

acquisition on the original learning under all conditions is satisfactorily uniform.

Retroactive Inhibition at Recall.—The results of the present experiment tend to bear out previous findings that the recall scores provide a more sensitive measure of RI than do relearning scores, *i.e.*, number of trials to reach a criterion of mastery. For that reason the major emphasis here will be upon the data concerned with the first 4 relearning trials.

Table III gives the mean correct anticipations on each of the first 4 relearning trials, together with the percentages of RI as determined by the usual formula, 100 (rest − work)/rest. On the first recall trial the difference between the mean number of correct anticipations after rest (condition A) and after 2 interpolated trials (condition B) is .88 ± .51 (t = 1.73). The differences between the means on the

TABLE III

MEAN CORRECT ANTICIPATIONS AND THE PERCENTAGES OF RI ON EACH OF THE FIRST 4 RELEARNING TRIALS

Condition	Relearning Trial							
	1		2		3		4	
	M	σM	M	σM	M	σM	M	σM
A	4.38	.59	5.63	.50	6.21	.56	6.92	.51
B	3.50	.44	5.71	.40	6.38	.49	7.08	.40
C	2.46	.48	3.83	.48	5.29	.55	6.21	.43
D	1.96	.45	4.50	.58	5.63	.60	6.42	.56
E	1.88	.53	4.08	.62	5.67	.55	6.33	.60
	Percent of RI							
B	20.1		−1.4		−2.7		−2.3	
C	43.8		31.9		14.8		10.3	
D	55.3		20.1		9.3		7.2	
E	57.1		27.5		8.7		8.5	

other work conditions and the rest condition are all reliable, well beyond the one percent level of confidence. The frequencies of anticipations on the first recall trial for conditions D and E are both sufficiently smaller than on condition B to be significant (t = 3.10 and 3.17 respectively).

Although RI is still obviously present on the second relearning trial, by the third relearning trial all differences are relatively non-significant, the largest difference, 1.09 ± .48 (t = 2.27), obtaining

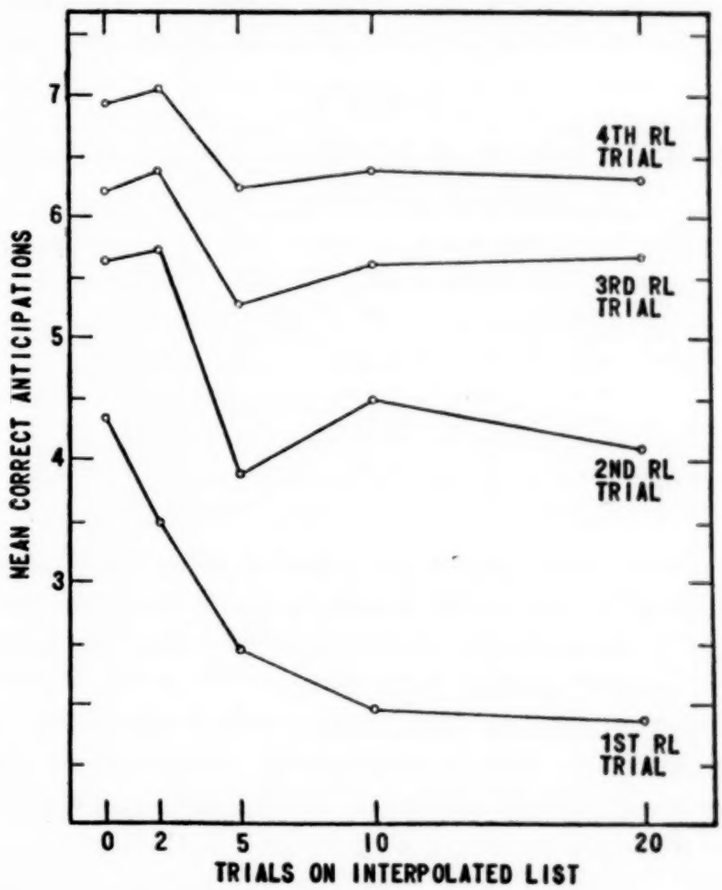


FIG. 1. Mean correct anticipations on the first 4 relearning trials with 0, 2, 5, 10, and 20 interpolated trials.

between conditions B and C. The trends in the dissipation of the RI are most clearly shown in the percent values of Table III.

Fig. 1 shows the trend of the inhibition in terms of the mean correct anticipations on the first 4 relearning trials as plotted against the 5 different degrees of interpolation. The general form of these curves is very similar to that obtained by Melton and Irwin (8). However, on the first cycle of their experiment, they found the greatest perseveration of RI to be coincident with 10 repetitions of the interpolated list, whereas Fig. 1 indicates that the greatest perseveration of RI falls at 5 repetitions of the interpolated list in the present experiment.

It should be pointed out that with the highest degree of interpolation used in this experiment (20 trials), RI is not less than with intermediate degrees of interpolated learning, nor is there any tendency for the RI to dissipate more rapidly with the highest degree of interpolation. These facts do not verify the Melton-Irwin results, although the 20 trials of interpolated learning in their experiment yielded only slightly more rapid dissipation than did their 10-trial condition. The possible causes of these discrepancies will be discussed later.

RI Measured by Relearning Scores.—There is little need to discuss the results under this category. Table IV gives the essential data.

TABLE IV

MEAN NUMBER OF TRIALS REQUIRED TO RELEARN THE ORIGINAL LISTS TO 1 PERFECT AND 2 SUCCESSIVE PERFECT TRIALS

Condition	1 Perfect		2 Perfect	
	M	σM	M	σM
A	8.96	1.31	10.54	1.56
B	9.13	1.04	10.67	1.14
C	9.79	1.12	11.50	1.35
D	10.00	1.44	11.33	1.70
E	10.00	1.38	12.38	1.70

Although the mean values show the trend in inhibition to be consistent with that exhibited by the recall scores, none of the differences between the mean trials is statistically significant, whether the criterion is one perfect or 2 successive perfect repetitions.

RI and Overt Inter-List Intrusions.—The measures of inter-list interference which have been presented up to this point have all been indirect. The most direct evidence available concerning the active competition of response systems is contained in the data on overt inter-list intrustions. Table V analyzes the intrusions as they occurred during the relearning of the original lists. This table shows the frequency of intrusions on each trial for the 4 work conditions. A

TABLE V

FREQUENCY OF SPECIFIC INTER-LIST INTRUSIONS OCCURRING DURING THE RELEARNING OF THE ORIGINAL LISTS

Condition	Relearning Trial					Total
	1	2	3	4	Remainder	
B	0	1	0	0	0	1
C	18	0	0	0	2	20
D	26	2	5	3	9	45
E	3	0	2	0	5	10

total of 76 intrusions occurred on all conditions, with by far the greatest total appearing on condition D, in which there were 10 interpolated trials. Not only are there more intrusions on the first 4 relearning trials under condition D, but there is also a greater frequency of intrusions occurring after these 4 trials.

Since nearly equal amounts of RI were found at the time of recall under conditions D and E (10 and 20 interpolated trials respectively), it would appear that intrusions cannot be used as independent indices of the absolute amounts of RI. In short, while intrusions indicate an active competition of response systems and hence some inhibition, interpretations of the gross frequencies as measures of RI must be made with considerable reservation. This discrepancy between RI and the number of overt intrusions after varying degrees of interpolation is entirely consistent with previous findings (**8**).

A few intrusions from the original list occurred during the learning of the interpolated list. On all conditions there was a total of 17 such intrusions, none for condition B, 2 for condition C, 4 for condition D and 11 for condition E. Compared with the frequencies of intrusions at recall, these frequencies are small.

Additional Data on Competition of Responses.—From the present experiment it is possible to derive rather precise data bearing on the concept of 'competition of responses' in paired-associates learning. Since overt intrusions are assumed to be the best observable index of competition of responses, it would seem important to examine the conditions under which these intrusions occur. In brief, the analysis consists of determining the number of reinforcements [2] which occurred on each pair of words of the original learning, then counting the number of reinforcements on the corresponding (A-C) pair of the interpolated list, and lastly, observing the response given at attempted recall.

First, it is necessary to verify the usual assumption that the number of reinforcements is an index of resistance to forgetting. This has been done by analyzing the pairs of condition A, in which no interpolated learning occurred. Since intra-list associations, if present, should be fairly equally distributed over all pairs by virtue of the different orders of presentation used, any pair which was reinforced once can be treated exactly the same as any other pair which received one reinforcement, and so on with 2, 3, and 4 reinforcements. Such a procedure ignores two conditions which might influence the analysis. (1) No note is taken of the trial on which a reinforcement occurred, that is, a reinforcement occurring on the second trial is assumed to be equivalent to a single reinforcement occurring on the fourth trial.

[2] Reinforcement is defined here as the giving of a correct response by the subject. One correct response will be called one reinforcement, 2 correct responses 2 reinforcements, etc.

(2) No cognizance is taken of any incorrect responses that are made. What influence these factors have on the analysis is not known, but it is assumed that their importance is not great enough to vitiate the procedure.

Condition A had a 20 min. rest period between learning and recall. Of the pairs receiving no reinforcements on the original learning, 10 percent became correct responses on the first recall trial. Of those pairs receiving one reinforcement, 46 percent were correct at recall; 52 percent were correct when they had been given 2 reinforcements; 86 percent were correct after 3 reinforcements, and 97 percent were correctly recalled after having received 4 reinforcements on the original learning. These data show quite clearly the relation between the frequency of reinforcement and strength of response-tendency, when strength is defined as resistance to forgetting.

In computing the number of reinforcements of the interpolated responses the various experimental conditions have been combined, since the variable assumed to be important is the number of reinforcements occurring and not the number of trials. Each pair of words in the original learning has been weighed against its corresponding pair in the interpolated learning by taking the difference between the number of reinforcements given each. That is, if the original pair had received 3 reinforcements and the interpolated item 5 reinforcements, a tabulation was made under the heading of '2 reinforcements in favor of the interpolated pair.' The total frequency for each possible category was calculated in this manner. Next, the difference between the number of reinforcements for the original and interpolated pairs was computed for all intrusions. Lastly, the frequency of intrusion for each category of relative reinforcement frequency was divided by the total cases in that category and multiplied by 100 to determine the percentage frequency of intrusion for each. The intrusions considered were only those on the first recall trial.

Table VI is a condensation of these data. It is limited here to the

TABLE VI

PERCENT OF INTRUSIONS OCCURRING FOR DIFFERENT RELATIVE REINFORCEMENT FREQUENCIES OF THE ORIGINAL AND INTERPOLATED PAIRS *

	Relative Frequency of Reinforcement							
	In Favor of IL					Equality	In Favor of OL	
	10–9	8–7	6–5	4–3	2–1	0	1–2	3–4
No. Cases	24	45	65	78	125	217	149	81
Intrusions	0	7	9	8	14	4	3	0
Percent Intrusions	0	16	14	10	11	2	2	0

* For complete explanation see text.

range of responses between '10 reinforcements in favor of the interpolated pair' and '4 reinforcements in favor of the original pair.' The number of cases outside this range is too small to admit their reliability. In addition, in Table VI the frequencies have been combined in adjacent categories.

It can be seen from Table VI that the great bulk of intrusions on the first relearning trial occur when the interpolated pair has received from one to 8 reinforcements more than the corresponding original pair. The percentage of intrusions occurring when the response-tendencies are of equal strength or when the original response-tendency is stronger is small. In addition, the data indicate that with great superiority in the number of reinforcements on the interpolated pair, the probability of an intrusion is decreased. Two intrusions did occur when the interpolated item had been reinforced 13 and 15 times respectively. These are not shown in the table.

There is one other source of data which gives additional evidence for conditions making for optimum frequency of intrusions. Forty-five percent of all intrusions occurred when the original pair corresponding to the intruding pair had received *zero* reinforcements; 14 percent occurred when the original pair had one reinforcement; 14 percent with 2 reinforcements; 12 percent with 3, and 14 percent with 4 reinforcements. It should be noted especially that nearly one-half of all intrusions appeared when the original pair, the supposedly competing pair, had never been reinforced.

IV. Discussion and Theoretical Implications

The results of this experiment confirm the study by McGeoch (3) and for the most part parallel the results obtained by Melton and Irwin (8). Since these *E*s utilized serial learning and nonsense syllables, whereas the present experiment employed the strict paired-associate technique with adjectives, the data suggest an extension of the generality of the relation between RI and the degree of interpolation. That is, RI increases with successive increases in the interpolated repetitions, but beyond moderate levels of interpolated learning little change is noted.

Melton and Irwin found a decrease in RI with 40 interpolated trials. In the present experiment, where 20 trials constituted the highest degree of interpolated learning, no comparable decrease was noted. However, in the Melton and Irwin experiment, 21 *S*s achieved the criterion of one perfect recitation in the 40 trials allotted, whereas in the present experiment only 17 *S*s reached this criterion. Thus, the mean level of learning achieved with maximal interpolation in the present experiment would coincide with the mean level of learning

achieved by Melton's and Irwin's *Ss* at some point between 20 and 40 interpolated trials. Since the decrease which Melton and Irwin report was found only with 40 repetitions of the interpolated list, a similar decrease in RI should not necessarily be expected in the present study. The fact that Melton and Irwin found the greatest perseveration of RI with 10 interpolated trials, whereas in the present experiment the greatest perseveration occured with 5 trials, probably can also be attributed to different levels of acquisition with equal numbers of trials.

The results of the present experiment have some bearing on current theoretical issues concerning RI. As previously mentioned, one of the purposes of this experiment was to provide a check on a theoretical curve (8) which has served as the foundation for the postulation of a two-factor theory of RI. This two-factor theory was a direct out-

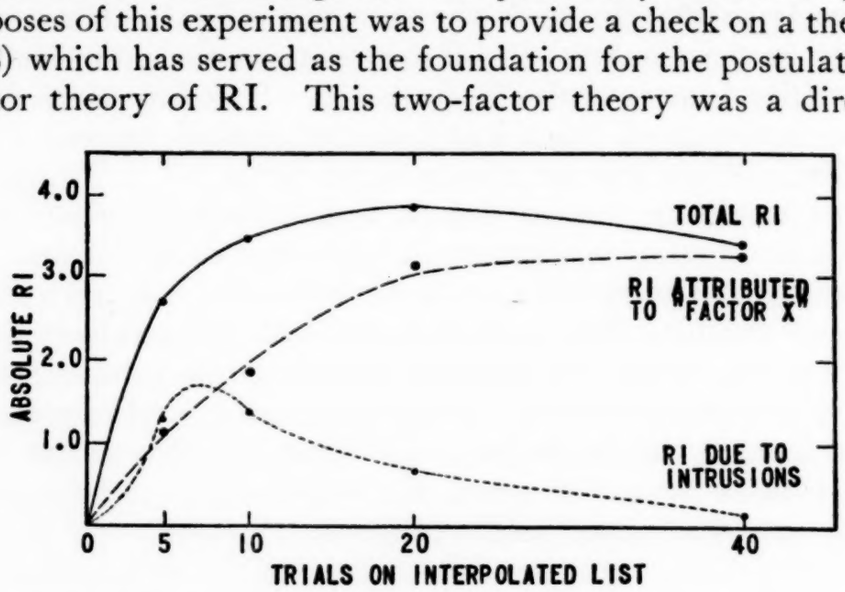


FIG. 2. Relationship between the absolute RI and the degree of learning of the interpolated list in an experiment performed by Melton and Irwin. The ordinate indicates the absolute decrement on the first recall trial as determined by the mean recall on the rest condition minus the mean recall on the work conditions. For further information see the discussion in the text. These curves were re-drawn from Melton's and Irwin's article (8).

growth of the Melton and Irwin study in which the degree of interpolated learning was varied. These writers discovered that on the first relearning trial the curve of RI as a function of the degree of interpolation was not coordinate with the curve of *overt* inter-list intrusions as a function of the same variable. Not only were the forms of the curves different, but the total overt intrusions could account for only a small portion of the total RI if it was assumed that each intrusion displaced what otherwise would have been a correct anticipation. Presumably a strict competition or reproductive inhibition theory of RI, as proposed by McGeoch (5), cannot incorporate these facts as stated. The competition theory allocates all of the measured inhibition to competition of covert and overt responses at recall, and, supposedly, the curves of RI and overt intrusions should at least show the same general form.

To account for the RI not attributable to the competition of responses, Melton and Irwin proposed an additional factor which they termed, non-committally, Factor X. The basis for the determination of the extent of Factor X is shown in Fig. 2, which has been re-drawn from their report. The top curve shows the empirical RI on the first recall trial. The lower curves shows the mean number of overt intrusions on the same trial when the empirical frequency of these intrusions is multiplied by two. This multiplier, two, has been used by Melton and Irwin to give weight to partial intrusions which occur at recall. The middle curve of the figure, the Factor X curve, indicates the inhibition that cannot be attributed to competition of responses when intrusions are used as the index of competition. They point out that while the shape of the Factor X curve could be changed

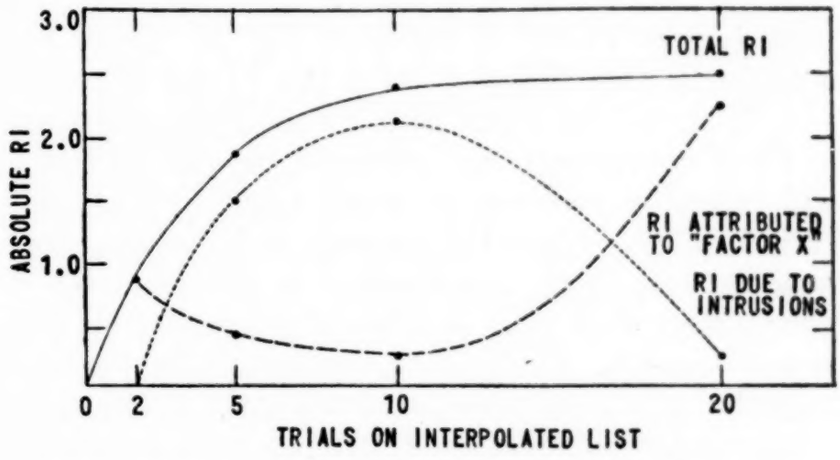


FIG. 3. Relationship between the absolute RI and the degree of learning of the interpolated list for the present experiment. The procedure for plotting these curves was the same as that used by Melton and Irwin in constructing the curves shown in Fig. 2. For complete information see the discussion in the text.

by using a different multiplier, the shape of the intrusion curve could never approximate the form of the curve of RI regardless of what multiplier is used.

Since the results of our experiment have closely paralleled the results of the Melton and Irwin study, an attempt was made to derive curves which would be comparable to those of Fig. 2. Fig. 3 shows the results of the present experiment when the procedure of these writers was duplicated. It is obvious at once that the Factor X curve of the present experiment differs considerably from the Factor X curve of Fig. 2. Furthermore, if Factor X represents the functioning of a single mechanism, it is incongruous with any known psychological function with which it might be legitimately compared. Before considering some possible causes for the discrepancy between the two experiments it is necessary to consider the interpretation that has been given to Factor X.

Tentatively, Factor X has been identified as a weakening or an unlearning of the original responses as a result of their non-reinforcement during the interpolated learning (**8, 9**). This weakening of the original responses should account for some of the inhibition at the time of recall since the relative response-tendencies of the original and interpolated lists will be changed. As has been recently pointed out by Melton and von Lackum (**9**), the unlearning theory is not contradictory to a strict competition-at-recall theory. The unlearning factor is postulated to emphasize the weakening of the original items which must change the relative response-tendencies of the original and interpolated items at recall.

Recent experiments (**6, 9**) have tended strongly to support unlearning as a fruitful construct. However, no measure of the unlearning component in the total RI at recall has yet been suggested other than that indicated by the Factor X curve of Fig. 2. There are several reasons for doubting the validity of using this Factor X curve for such purpose. (a) It should be noted that if the interpolated learning in the Melton-Irwin experiment was increased beyond 40 trials, the RI curve should probably continue to come down toward the abscissa, indicating still less RI. Since the number of intrusions would probably decrease likewise, their contribution to the total RI would be practically zero, since even with 40 trials they account for only a very small portion of the RI. Under such conditions the Factor X curve would necessarily have to decrease, and this state of affairs for such a factor as unlearning could hardly be justified since it would indicate first a weakening and then a strengthening of the original words as the interpolated learning continued.

(b) Fig. 3 showed that the Factor X curve of the present experiment exhibited peculiar characteristics. This casts some doubt on the validity of identifying Factor X with a single postulated mechanism, such as unlearning.

(c) It was shown in the present experiment that nearly half of all instrusions occurred when the original response had received zero reinforcements and hence was 'unavailable' at recall. The curves of Figs. 2 and 3 are based on the obviously false assumption that each intrusion takes the place of what otherwise would have been a correct response.[3]

(d) Some question may be raised as to the legitimacy of assuming that the partial-complete intrusion ratio remains constant with vary-

[3] It was shown in the item analysis that with no interpolation (condition A), responses receiving no reinforcements on the original learning were correctly recalled 10 percent of the time. However, with interpolated learning only 3 percent of these pairs were correctly recalled even though the corresponding interpolated pair had received no reinforcements. No intrusions took place under such conditions.

ing degrees of interpolated learning. This assumption was basic to the use of the constant multiplier, two, for all degrees of mastery of interpolated material. Furthermore, since implicit intrusions probably do occur, this form of interference should be considered on the same dimension with partial and complete intrusions. By an implicit intrusion we may mean at least two conditions, both reported by *Ss*: (1) two responses which are of near equal strength so that the *S* does not have time or cannot make an overt choice,[4] or, (2) an interpolated response which is so strong that it 'dominates' the response from the original learning. This latter form of implicit or covert intrusion is suggested by *Ss* who report, "I thought of a response but I knew it was wrong, though I didn't have time to think of the correct response."

If it be assumed that the covert-partial-overt intrusion ratio changes with varying degrees of interpolation, these varying types of intrusions could account for all of the measured RI. Such a theory can include the unlearning or weakening of the original responses since this weakening will change the relative strengths of response-tendencies of the original and interpolated items, and consequently the intrusion ratios would be changed.

Two main problems must be met before even a crude outline of a theory based on changing covert-partial-overt intrusion ratios can be stated: (a) a statement of the change occurring in the ratio as changes occur in the relative reinforcement frequency of the original and interpolated responses, and (b) the amont of influence which unlearning has in the weakening of the original responses.

With respect to the first problem, our objective data and the subjective reports are suggestive of an inverse relation between covert and overt intrusions. Table VI showed that the greatest percentage of intrusions appeared when the interpolated response was from one to 8 reinforcements stronger than corresponding original response. With greater disparity in relative frequency of reinforcements, intrusions were reduced to a minimum, though RI changed but little. The subjective reports (as indicated above) suggest that interference in the form of covert intrusions occurred when the interpolated response had a great many more reinforcements than the original response, and that actual blocking occurred when the two responses were of nearly equal strength as determined by the objective frequency of reinforcements. There is a precedent for suggesting such a relationship between the overt and covert intrusions. Siipola (11) has shown essentially this ratio between intrusions in a simple manual habit, and she suggests that the two forms, overt and

[4] Other experimenters have suggested this form of interference, usually called 'blocking' (*cf.* 1, 8, 10).

partial errors, are complementary. We carry the relationship one step further and say that implicit and partial errors combined may be inversely related to overt intrusions.

The second problem, the influence of unlearning, is more difficult to handle. However, the above statement of the inverse relation between covert and overt intrusions requires that the influence of unlearning be relatively constant after a 'few' trials on the interpolated list. It seems logical that if a weakening of the original responses is occurring during the interpolated learning, some indication of it should be available from the objective records of the learning of the interpolated list. There are two such measurements, the appearance of intrusions from the first list during the learning of the second list and associative inhibition (retardation in learning) occurring during the acquisition of the second list. This associative inhibition is usually quite transitory (**7, 9**), its extent apparently depending primarily upon the degree of learning of the first list and the similarity between the two lists. Melton and von Lackum (**9**) have suggested the possibility that the amount of unlearning might be indicated by the amount of associative inhibition present during the learning of the interpolated list. Since such inhibition is usually transitory, though subject to variation by the above mentioned conditions, the rather indefinite word, 'few,' has been used to indicate the number of trials on the interpolated list which must occur before unlearning ceases to increase.[5]

Lastly, it should be remembered that Melton and Irwin used nonsense syllables and serial learning, and although RI as a function of degree of interpolation was very similar to the results of the present experiment, the greater complexity of serial learning must to some extent modify the transfer of principles from one form of learning to the other. What we have tried to do in the present discussion is to indicate the limitations of the characteristics initially given to the unlearning construct and to make some guesses as to the lack of validity in the assumptions used in describing these characteristics.

V. Summary

Twenty-four *S*s learned lists of 10 paired adjectives by the anticipation method for 5 trials, rested 20 min. or learned a second list for

[5] A comparison of recent experiments (**6, 9**) suggests the possibility that the amount of unlearning of the original responses is less with paired associates than with serial learning. If unlearning is to be identified with associative inhibition there should likewise be less associative inhibition with paired associates than with serial lists. Although some data (**6, 9**) suggest that this may be true, the difficulty of comparing the interfering or similarity factor of the A–B, A–C paradigm in paired associates learning with the similarity factors in serial learning makes such inter-experiment comparisons of questionable value.

2, 5, 10 or 20 trials, and then relearned the original list to a criterion of 2 successive errorless trials. The results tend to support the following generalizations:

1. With increasing amounts of interpolation up to 10 repetitions of the interpolated list, RI as measured by recall scores is a function of the degree of interpolated learning. After 10 trials on an interpolated list, an additional 10 trials produces no further increase in RI.

2. No significant amounts of RI were present when measured by trials to reach one perfect and 2 successive perfect repetitions.

3. The data from overt inter-list intrusions show that the greatest frequency of these intrusions is with 10 interpolated learning trials, less with 5 and 20 interpolated trials, and practically zero with 2 interpolated trials.

4. An item analysis of the relative response-tendencies of the original and interpolated responses shows that nearly all the overt intrusions occur when the interpolated response is somewhat stronger than the original response, with few intrusions occurring when the items are of near equal strength, or when the original item is stronger, or the interpolated item much stronger.

5. There is essential agreement between the results of this experiment and related experiments by McGeoch (**3**) and Melton and Irwin (**8**).

The following points were brought out in the theoretical discussion:

1. To identify Factor X of the two-factor theory of RI as wholly an unlearning or weakening of the original responses makes the unlearning factor possess incongruous psychological properties, at least in the present experiment.

2. The argument was put forward that the overt-partial-covert intrusion ratio is not constant with varying degrees of interpolation.

3. The following hypotheses were briefly formulated: (a) an inverse relation obtains between overt and covert intrusions, and (b) the amount of unlearning of the original list is constant after a 'few' trials on the interpolated list.

(Manuscript received October 5, 1942)

References

1. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
2. Lahey, M. F. L. Retroactive inhibition as a function of age, intelligence, and the duration of the interpolated activity. *Cath. Univ. Amer. Educ. Res. Monogr.*, 1937, **10**, No. 2, 93 pp.
3. McGeoch, J. A. The influence of degree of interpolated learning upon retroactive inhibition. *Amer. J. Psychol.*, 1932, **44**, 695–708.
4. McGeoch, J. A. Studies in retroactive inhibition. VII. Retroactive inhibition as a function of the length and frequency of presentation of the interpolated lists. *J. exp. Psychol.*, 1936, **19**, 674–693.

5. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green & Co., 1942. Pp. xvii + 633.
6. McGeoch, J. A., & Underwood, B. J. Tests of the two-factor theory of retroactive inhibition. *J. exp. Psychol.*, **1942, 32,** 1–16.
7. Melton, A. W. Proactive inhibition in serial verbal learning as a function of the degree of learning of the first material. *Psychol. Bull.*, 1936, **33,** 765.
8. Melton, A. W., & Irwin, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53,** 173–203.
9. Melton, A. W., & von Lackum, W. J. Retroactive and proactive inhibition in retention: evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54,** 157–173.
10. Sears, R. R., & Hovland, C. I. Experiments on motor conflict. II. Determination of mode of resolution by comparative strengths of conflicting responses. *J. exp. Psychol.*, **1941, 28,** 280–286.
11. Siipola, E. M. Implicit or partial reversion-errors; a technique of measurement and its relation to other measures of transfer. *J. exp. Psychol.*, 1940, **26,** 53–73.
12. Twining, P. E. The relative importance of intervening activity and lapse of time in the production of forgetting. *J. exp. Psychol.*, 1940, **26,** 483–501.

SENSITIZATION AND ASSOCIATION IN EYELID CONDITIONING[1]

BY DAVID A. GRANT

University of Wisconsin

Introduction

In a previous study (**4**), reliable increases in the frequency of eyelid responses to light were found to occur when the cornea was repeatedly stimulated by an air-puff, and these changes were retained for an interval greater than one week. The frequency of responses elicited by auditory stimuli followed similar trends but these trends were not as reliable. In general, the changes in frequency of responses to the light and sound stimuli resembled conditioning, but the responses could not have been conditioned responses, in the ordinary meaning of the term, because the visual and auditory conditioned stimuli were never *paired* with the puff stimuli during reinforcement. Such responses developed following repetitive stimulation with an unconditioned stimulus have been called 'pseudo-conditioned' responses (**7**).

In a control procedure of the earlier experiment subjects simply sat and maintained their visual fixation for the same length of time that members of the experimental group had sat and fixated while receiving the puff-series. An increase in response frequency similar to that on the first day of the experimental procedure was found in the control group. This group was not tested for retention.

It was especially significant that the eyelid responses which increased in frequency in both the experimental and control groups were of relatively long latency, 170 to 450 msec. This is the usual latency range of conditioned eyelid responses (**10**, p. 231) and indicates a similarity between the sensitized and the conditioned eyelid responses in latency range as well as in frequency increment. Unfortunately, no orthodox conditioning procedure was used in the earlier study so that detailed comparison of sensitized and conditioned eyelid responses was impossible.

The similarity between the sensitized eyelid responses of the

[1] This paper is based upon a dissertation submitted to the Department of Psychology and the Committee on Graduate Study of Stanford University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer wishes to thank Professor C. P. Stone, Professor Quinn McNemar, and especially Professor E. R. Hilgard for guidance and encouragement during the course of this investigation.

earlier study and previously reported conditioned eyelid responses directed interest in the present investigation. The present study consists of an experimental comparison of sensitized and conditioned eyelid responses developed under the same general experimental conditions.

Procedure

The apparatus and experimental conditions were essentially the same as those of the previous study (4). The conditioned stimulus was an increase in brightness of a 10 cm. square milk-glass screen from about .02 to about 300 millilamberts, as measured with a Macbeth Illuminometer. The duration of the light stimulus was about one sec. The subject was instructed to fixate a small black cross on the milk-glass screen, except during rest periods. The unconditioned stimulus was a puff of air to the left cornea. When the conditioned and unconditioned stimuli were presented together, the puff was delivered about 400 msec. after the onset of the light. Trials were given at random intervals ranging from 25 to 45 sec.

The design and procedure of the present experiment are outlined in Table 1 and are briefly summarized below.

TABLE 1

Experimental Procedure

Procedure	Group I Pseudo-conditioning (N = 32[1])	Group II Conditioning (N = 31[2])	Group III Fixation (N = 31[3])	Group IV Control (N = 30)
Day 1				
Pre-test.....	4 light	4 light	4 light	—
Reinforcement...	40 puff	40 light-puff	fixation	40 puff
Post-test.....	4 light	4 light	4 light	4 light
Day 2				
	Same as Day 1 for Groups I, II, and III.			—
Day 3				
Extinction.....	8 light (jet) 8 light (no jet) 8 light (jet)	8 light (jet) 8 light (no jet) 8 light (jet)	24 light (no jet)	—

[1] Three of these failed to appear for the third day of the experiment.
[2] Two of these failed to appear for the third day of the experiment.
[3] One of these failed to appear for the second and third days of the experiment.

Group I.—The pseudo-conditioning procedure was used with a group of 32 subjects. It consisted of a 40-trial puff-series, preceded and followed by pre- and post-tests, each made up of four trials with the light alone. Light and puff were never presented together. The procedure was the same on each of the first two days.

Group II.—The conditioning procedure was used with a group of 31 subjects. The pre-test and post-test with the light alone were given as in the pseudo-conditioning procedure, but the light and puff were presented together during the 40-trial reinforcement series on each of the first two days of the experiment.

Group III.—The fixation procedure was used with a group of 31 subjects. It was the same as the pseudo-conditioning procedure except that *no* puff stimulation was given. During the

time intervening between the pre-test and the post-test with the light alone, the subject simply sat and looked at the small fixation cross on the milk-glass screen. No stimulation was given during this interval.

Group IV.—A control group of 30 subjects was run through the first day procedure of the pseudo-conditioning group except that the 4-trial pre-test with the light alone was omitted. (It was conjectured that because of the initial series of light stimuli, some sort of bond with the air-puff might have been established in the pseudo-conditioning group. The control group served to check on this possibility.)

On the third day of the experiment, subjects of the pseudo-conditioning, conditioning, and fixation groups were given an extinction series of 24 light stimuli. To check on the influence of expectation of the puff during extinction the air-jet, which served as a means of presenting the unconditioned stimulus, was removed for the second series of eight extinction trials in the pseudo-conditioning and conditioning groups. The jet was *never* present for members of the fixation group.

Results

1. *Latency.*—Under the conditions described above, there appeared in all four groups of subjects eyelid reactions similar to those found in orthodox eyelid conditioning experiments. Examination of the latencies of these responses provided a basis for classifying them as reflex or non-reflex in origin.

The latencies of all eyelid reactions evoked by the light alone (on the pre- and post-tests and extinction series) are plotted in Fig. 1 for each experimental group.

The first impression obtained from Fig. 1 is that pseudo-conditioning, conditioning, and visual fixation have resulted in eyelid responses with similar latency distributions. All of the latency distributions are bimodal. For each procedure the first mode—at 130 msec.[2]—is made up of unconditioned reflexes evoked by the light. The second modes are at 200 msec. in Groups I, III, and IV, and at 180 msec. in Group II. Beyond the second mode, all of the distributions tend to flatten out.

For purposes of the present study, all eyelid reactions to light with latencies of 170 to 450 msec. inclusive were defined as 'pseudo-conditioned responses' (PCR's), 'conditioned responses' (CR's), or 'fixation responses' (FR's), according to the reinforcement procedure which had been used.[3]

A careful comparison of the latency distributions of Fig. 1 for the range from 170 to 450 msec. reveals that the CR's clearly differ from the sensitized responses in having a smaller mode at a lower latency.

[2] The latencies are as determined from the records and are not corrected for the latency of recording devices. The absolute latency is always approximately 50 msec. less than reported.

[3] The latency of 160 msec., delimiting the PCR's, CR's, and FR's from the unconditioned reflexes to the light stimulus, was chosen because the frequency of responses between the modes in each distribution was a minimum at this latency. The upper limit of 450 msec. was chosen because eyelid CR's to light may have measurable latencies that high when the puff is presented 400 msec. after the onset of the light. CR's of greater latency are obscured by reflexes to the puff. The same limit was used with the sensitization groups in order that the treatment of results might be comparable.

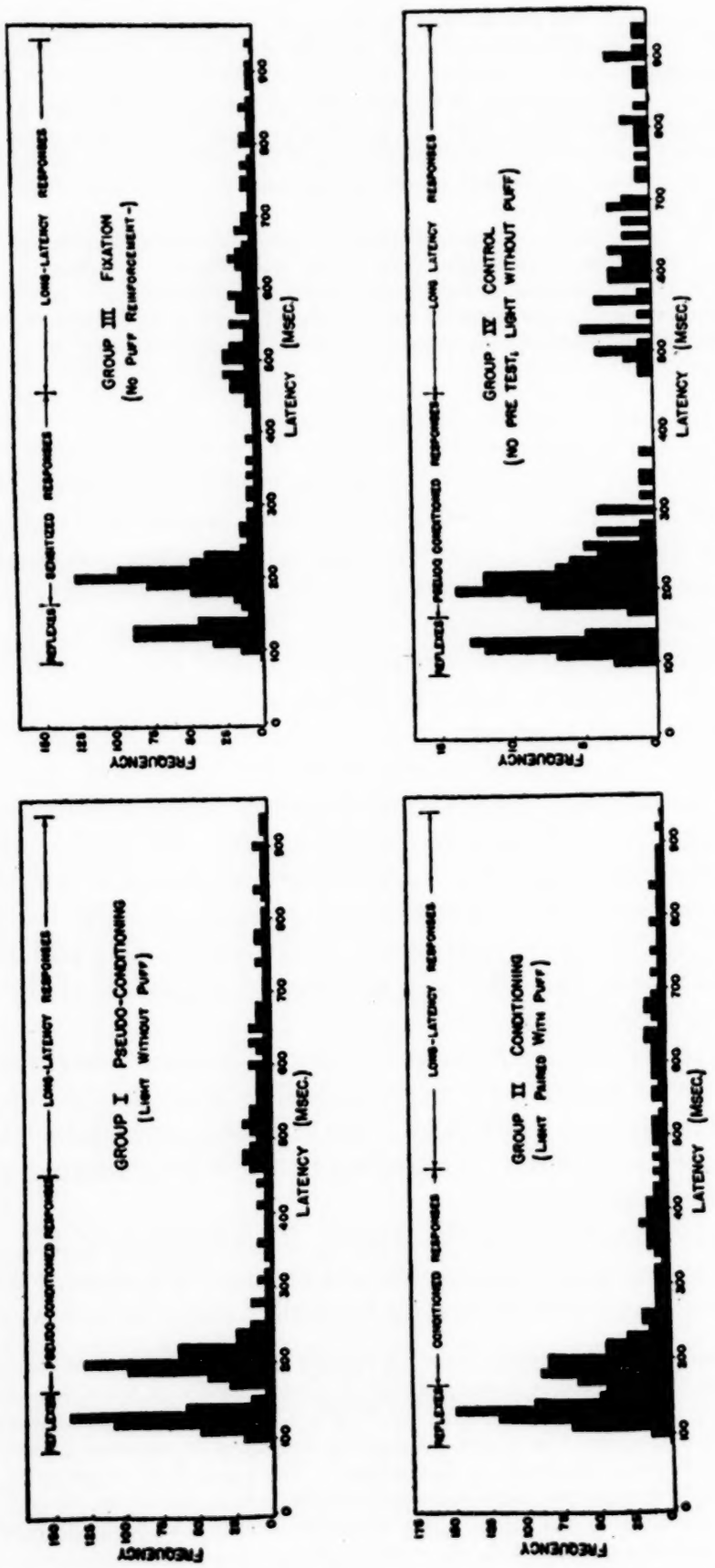


FIG. 1. Distributions of the latencies of eyelid responses evoked by the light stimulus alone.

Chi-square (3) was found to be 91 between groups I and II with 8 degrees of freedom, 99 between groups II and III with 19 degrees of freedom, and 34 between groups I and III with 17 degrees of freedom. The large values of chi-square indicate systematic (non-random) differences between the latency distributions [4] of PCR's, CR's and FR's.

2. *Frequency Results.*—a. Trends following reinforcement and non-reinforcement.—The mean frequencies of PCR's, CR's and FR's evoked by the light stimulus during the different test series of the experiment are given in Table 2 along with their SD's. The 24 trials

TABLE 2

FREQUENCY OF PSEUDO-CONDITIONED, CONDITIONED, AND FIXATION RESPONSES

Group		Day 1		Day 2		Day 3					
		Pre	Post	Pre	Post	Successive Four-Trial Blocks					
						1	2	3	4	5	6
I. Pseudo-conditioning...	Mean	1.0	2.8	1.5	3.0	1.8	1.7	2.0	2.5	2.0	2.4
	SD	1.4	1.2	1.2	1.1	1.6	1.5	1.3	1.5	1.4	1.5
II. Conditioning.....	Mean	1.1	2.9	2.6	3.0	2.5	2.4	2.3	2.3	2.5	2.3
	SD	1.3	1.1	1.3	1.1	1.0	1.3	1.4	1.4	1.4	1.5
III. Fixation.....	Mean	1.2	2.8	1.7	2.7	1.9	2.2	2.1	2.4	2.6	2.6
	SD	1.5	1.3	1.6	1.2	1.4	1.5	1.2	1.4	1.3	1.4
IV. Control Pseudo-conditioning.....	Mean		3.1								
	SD		0.9								

of day 3 have been broken down into six successive four-trial blocks. The frequency data from Table 2 are presented graphically in Fig. 2 to facilitate comparison of results.

Examination of Table 2 and Fig. 2 shows that on the pre-test of day 1 there were no important differences among the three experimental groups. Increases in frequency from pre-tests to post-tests are readily apparent in all groups for day 1 and day 2. The control-group frequency is comparable to the pseudo-conditioning-group frequency on the post-test of day 1. This indicates that the pre-test with light is not crucial to the development of PCR's in Group I. Overnight decreases appear in all experimental groups between days 1 and 2 and between days 2 and 3.

The statistical significance of each daily increase and of each overnight decrease is indicated by its critical ratio in Table 3. The

[4] The group IV distribution was omitted from the comparisons because it was based upon only one-tenth the amount of data involved in the other comparison.

increases following pseudo-conditioning, conditioning, and fixation on day 1 are all highly significant. The day 2 increases of PCR's and FR's are also significant. Overnight decreases are statistically significant for the PCR's and FR's but not for the CR's.

b. Absence of extinction.—No regular trends in frequency are apparent during the six 4-trial blocks on day 3. The most striking aspect of the day 3 frequency data is the complete absence of ex-

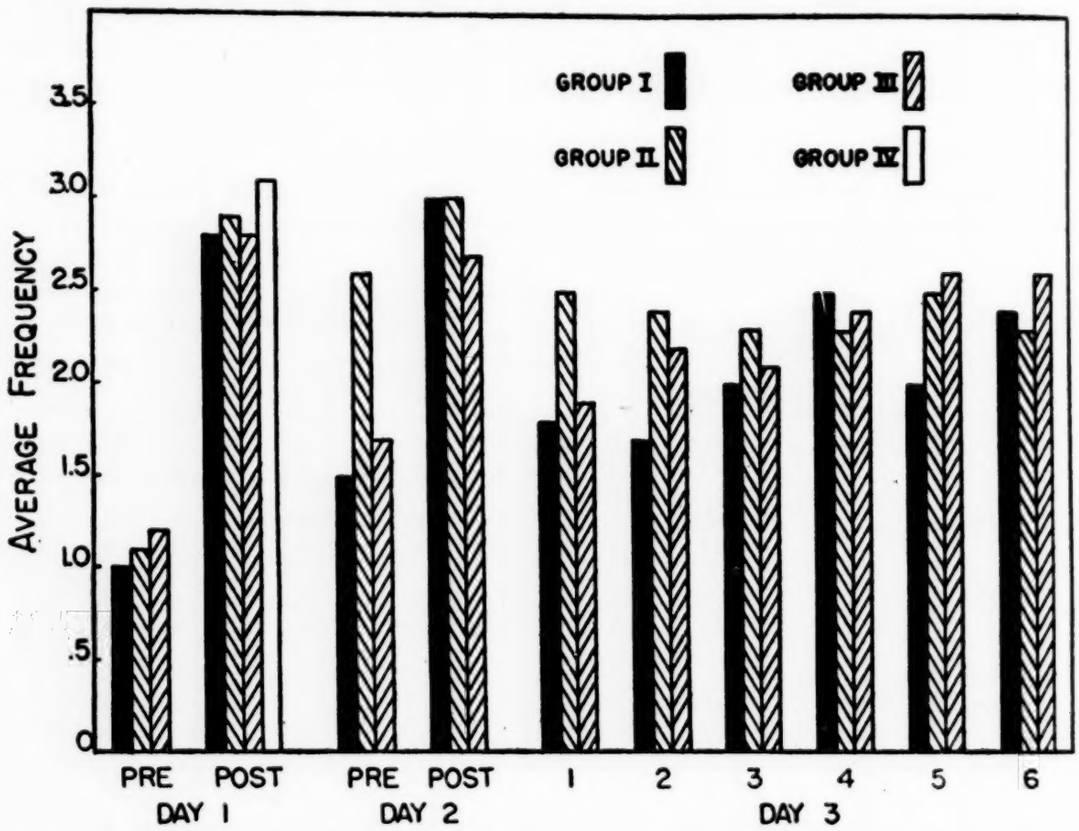


FIG. 2. Average frequencies of pseudo-conditioned, conditioned, and fixation responses evoked by the light stimulus alone. The solid black columns represent the frequency of PCR's, Group I; the columns cross-hatched down from left to right represent the frequency of CR's, Group II; the columns cross-hatched up from left to right represent the frequency of FR's, Group III; and the outlined column represents the frequency of PCR's in the control group, Group IV.

tinction trends. This is in agreement with previous results on the pseudo-conditioned eyelid response (4), and with the resistance to extinction which has been reported for pseudo-conditioning of other types of responses (19, 7, 8). However, the orthodox conditioned eyelid response, developed in human subjects,[5] usually (10, 17, 16, 13), but not always (18), extinguishes rapidly. The absence of extinction probably resulted from the relaxing instructions used in the present study or from the intensity of the light stimulus. Further data are needed before the crucial factors are determined.

[5] Hilgard and Marquis (11) found no extinction of conditioned eyelid responses in monkeys.

In the pseudo-conditioning and conditioning groups the air-jet was removed after the second 4-trial block and replaced after the fourth 4-trial block as a check on the influence of expectancy. Following the removal of the jet, contrary to the anticipated results, the frequency of PCR's *increased* reliably from the second to the fourth block, and following reintroduction of the jet the frequency of PCR's *decreased* reliably. Increased expectation of the puff following reintroduction of the jet was reported by 23 out of 29 subjects, but the increased expectation did not serve to effect a corresponding increase in responsiveness.

Removal and reintroduction of the air-jet had the opposite effect upon the frequency of CR's, although the changes were not statistically significant. The direction of the changes were in line with Humphrey's 'expectancy hypothesis' (**13**).

TABLE 3

CRITICAL RATIOS FOR DAILY INCREASES AND OVERNIGHT DECREASES IN MEAN FREQUENCY

Group	Day 1 Increase	Overnight Decrease Day 1 to Day 2	Day 2 Increase	Overnight Decrease Day 2 to Day 3
I. Pseudo-conditioning	6.91	5.44	6.69	4.09
II. Conditioning	7.93	1.60	1.84	1.91
III. Fixation	5.25	2.77	3.49	2.98

c. Overnight retention.—The question of whether there is significant overnight retention of the frequency increases is of importance in evaluating the stability of the changes brought about by the different experimental procedures. It will be noted in Fig. 2 that in spite of the overnight decreases in all groups, the mean frequencies on the pre-test of day 2 and on the successive 4-trial blocks of day 3 remain greater than the average frequencies on the pre-test of day 1. With minor exceptions these differences were statistically significant as shown by their critical ratios in Table 4. This demonstrates reliable overnight retention of PCR's and CR's, and near-reliable retention of FR's.

d. Inter-group frequency differences.—Inter-group differences were to be expected because of the different types of reinforcement. However, differences were found only following overnight retention periods. On the pre-test of day 2, the frequency of CR's reliably (critical ratio = 3.14) exceeded the frequency of PCR's and nearly reliably (critical ratio = 2.30) exceeded the frequency of FR's. Suggestive, but less reliable, differences between the same groups appeared on the first 4-trial block of day 3.

The only outstanding inter-group difference in frequency *changes*

was the reliable (critical ratio = 3.04) difference between the decrease of CR's and the decrease of PCR's during the day 1 to day 2 overnight retention interval. This indicates a tendency for greater overnight retention of CR's as compared to PCR's.

3. *Amplitude and Duration Results.*—a. Amplitude results.—The changes in amplitude of PCR's, CR's, and FR's corresponded in almost all instances with the frequency trends, except on day 3. Here slight but consistent extinction trends were noted. Throughout the experiment the amplitude of CR's and FR's tended to exceed that of PCR's. However, no reliable changes and no reliable inter-group differences in amplitude could be shown to have occurred as a result of the different experimental procedures.

b. Duration results.—In order for the conditioned eyelid response to be effective as an avoidance reaction, complete or partial

TABLE 4

CRITICAL RATIOS FOR THE DIFFERENCES BETWEEN THE MEAN FREQUENCY ON THE PRE-TEST OF DAY 1 AND THE MEAN FREQUENCIES FOLLOWING OVERNIGHT RETENTION PERIODS

Group	Day 2	Day 3					
	Pre-test	Successive Four-Trial Blocks					
		1	2	3	4	5	6
I. Pseudo-conditioning.....	3.02	3.39	2.07	3.34	4.71	3.42	4.31
II. Conditioning.....	5.63	5.38	4.14	3.95	3.50	4.19	3.95
III. Fixation.....	2.19	2.06	3.25	2.87	3.37	4.45	3.61

closure must persist until after the puff has been given. Since the sensitized responses of the pseudo-conditioning and fixation groups did not serve such a function, it was thought that they might be of shorter duration [6] than the CR's.

The conjecture that CR's would tend to be of longer duration than PCR's and FR's was borne out by the average values. However, on none of the test-series were the inter-group differences statistically significant, although on some tests they closely approached significance. Otherwise, the duration data corresponded quite closely to the trends of the frequency data, showing increments following reinforcement and decrements following periods of non-reinforcement.

[6] For the purposes of this study duration was defined as the length of time that the amplitude of eyelid closure exceeded three mm. Duration was measured to the nearest hundredth of a second along an artificial baseline three mm. above the original eyelid baseline. The artificial baseline was necessary because after closure, the return of the eyelid to its original baseline is often asymptotic. The measure used proved to be fairly reliable. Spearman-Brown corrected reliability coefficients between the average duration of odd and even responses were .86, .94, .91, and .84 for groups I, II, III and IV, respectively.

Discussion

1. *Implications for Conditioning Theory.*—The most fundamental finding of this study is the great similarity of eyelid responses developed by associative (conditioning) and non-associative [7] (sensitization) techniques. The results give the impression that the observed differences were superimposed upon an underlying identity in response. The findings of this study are thus in complete agreement with previous data on the sensitization of the eyelid response (**4**). It seems possible to conclude, therefore, that many of the changes in eyelid conditioning are not primarily dependent upon the specific circumstances of the association of the conditioned and unconditioned stimuli during reinforcement.

It would be a mistake, however, to assume that the *pairing* of the conditioned and unconditioned stimuli has no effect whatever upon the eyelid response. Associative factors do play an important role. The CR's showed a different distribution of latencies than the sensitized responses; the CR's tended to be of greater duration and tended to be better retained overnight. These characteristics of the CR's favor an interpretation in terms of puff-avoidance. The CR's to the light, beginning prior to the puff and having sufficient amplitude and duration, serve adequately to protect the eye from the air-puff. Such a protective function could not be served by the sensitized responses, since no puff-avoidance was possible.

On their initial appearance the CR's and the sensitized eyelid responses seem to be basically identical. The CR's are later modified to conform to the discriminated features of the experiment. Probably these adaptive modifications are the only aspects of eyelid conditioning which should be considered associative learning. They may take place according to any or all of the commonly accepted principles of reinforcement (**12**).

The overlaying of associative changes on more primitive sensitized reactions is illustrative of a general process Dodge (**2**) has referred to as the 'epicritic modification of protopraxic adjustments.' The sensitized eyelid response is protopraxic, crude and relatively non-discriminative; the conditioned eyelid response is epicritic, relatively precise and discriminative, closely attuned to the specific situation.

2. *Sensitization and Association in Eyelid Conditioning.*—In the present experiment non-associative (sensitizing) features are very likely of greater importance relative to associative features than has

[7] It cannot be claimed that pseudo-conditioning involves *no* associative reinforcement. Wickens and Wickens (**23**) have some evidence to support their suggestion (**22**) that pseudo-conditioning is dependent upon a kind of redintegration. Mediation (**20, 21, 15, 14, 1**) is another possibility. Other studies (**7, 9, 6, 5**) have shown that spatial contiguity is a factor in the generalization of pseudo-conditioning.

been the case in many previously reported experiments. The absence (or delay) of extinction of the CR's suggests that this is the case. Emphasis on sensitization, however, serves only to dramatize the precautions which are necessary if eyelid conditioning experiments are to be interpreted as associative learning.

Previously reported experiments on eyelid conditioning apparently include in their results an undetermined proportion of changes resulting from unsuspected non-associative factors, in addition to the changes brought about by the more strictly associative aspects of the reinforcement. Further experimentation is required to establish the specific circumstances leading to the prevalence of sensitization or association within the same broad experimental arrangements. In the absence of knowledge of these circumstances, a necessary restriction of interpretation of data should be made. Before any eyelid conditioning results may be attributed to the association of the conditioned and unconditioned stimuli during reinforcement, suitable control procedures must be used to define the nature and extent of sensitization resulting from non-associative aspects of the experimental situation.

Summary

In order to study the effects of sensitization and association in eyelid conditioning, eyelid responses were photographically recorded during: (a) pseudo-conditioning (32 subjects), in which the conditioned stimulus was not presented with the unconditioned stimulus, (b) conditioning (31 subjects), in which the conditioned and unconditioned stimuli were paired in the usual manner, and (c) a visual fixation procedure (31 subjects), in which the unconditioned stimulus was not presented at all. The conditioned stimulus was an increase in illumination; the unconditioned stimulus was a puff of air to the cornea.

The basic conditioning procedure consisted of a 40-trial light-puff reinforcement series, preceded and followed by test-trials with the light alone. The procedure was the same on each of the first two days. The pseudo-conditioning procedure was like the conditioning procedure except that the reinforcement trials consisted of 40 presentations of the puff alone unaccompanied by the light. The fixation procedure was like the pseudo-conditioning procedure except that the air-puff series were omitted. During the time intervening between the initial and final tests with the light alone the subject simply sat and looked at a small fixation cross.

On the third day of the experiment all the subjects of the above-mentioned groups were given an extinction series of 24 light stimuli.

As a control on expectation of the air-puff, the air-jet was removed for one-third of the extinction trials for the pseudo-conditioning and conditioning groups; it was never present for the fixation group.

A control group of 30 subjects was run through the first day procedure of the pseudo-conditioning group except that the pre-test with the light was omitted. This checked on the possibility that the pre-test might have resulted in some association between the light and puff in the pseudo-conditioning group.

The results of the investigation may be summarized as follows:

1. Clear-cut and approximately equal increases in the frequency of eyelid responses were found following pseudo-conditioning, conditioning, and visual fixation procedures.

2. Essential similarity of results with the control and pseudo-conditioning groups demonstrated that associations between the initial light stimuli and subsequent puff stimuli were not important in determining the pseudo-conditioning results.

3. Although there were overnight decreases in response frequency, significant overnight retention was found in all groups, particularly in the conditioning group.

4. There was no extinction trend in the response frequency of any group during a series of 24 unreinforced presentations of the conditioned stimuli.

5. Orthodox conditioned responses differed from the sensitized eyelid reactions in that they: (a) tended to be of greater duration; (b) had a lower modal latency; (c) tended to show greater overnight retention.

6. The only statistically reliable difference between the responses of the pseudo-conditioning and fixation groups was in their distributions of latencies. Thus the role of the puff in eyelid pseudo-conditioning remains to be determined by further experimentation.

The striking similarity of eyelid responses following sensitization (pseudo-conditioning and fixation) and association (orthodox conditioning) suggests that conditioned and sensitized eyelid responses are fundamentally identical in earlier stages of reinforcement. It is suggested that in conditioning the responses are gradually modified through associative reinforcement so that they become adaptively related to the specific stimulus sequences.

(Manuscript received August 18, 1942)

References

1. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
2. Dodge, R. Protopraxic and epicritic stratification of human adjustments. *Amer. J. Psychol., Washburn Commemorative Volume*, 1927, 145–157.

3. FISHER, R. A. *Statistical methods for research workers.* (7th ed.) London: Oliver and Boyd, 1938.
4. GRANT, D. A. The pseudo-conditioned eyelid response. *J. exp. Psychol.*, 1943, **32**, 139–149.
5. GRANT, D. A., & DITTMER, D. G. A tactile generalization gradient for a pseudo-conditioned response. *J. exp. Psychol.*, 1940, **26**, 404–411.
6. GRANT, D. A., & MEYER, H. I. The formation of generalized response sets during repeated electric shock stimulation. *J. gen. Psychol.*, 1941, **24**, 21–38.
7. GRETHER, W. F. Pseudo-conditioning without paired stimulation encountered in attempted backward conditioning. *J. comp. Psychol.*, 1938, **25**, 91–96.
8. HARLOW, H. F. Forward conditioning, backward conditioning, and pseudo-conditioning in the goldfish. *J. genet. Psychol.*, 1939, **55**, 49–58.
9. HARLOW, H. F., & TOLTZIEN, F. Formation of pseudo-conditioned responses in the cat. *J. gen. Psychol.*, 1940, **23**, 367–375.
10. HILGARD, E. R., & CAMPBELL, A. A. The course of acquisition and retention of conditioned eyelid responses in man. *J. exp. Psychol.*, 1936, **19**, 227–247.
11. HILGARD, E. R., & MARQUIS, D. G. Conditioned eyelid responses in monkeys, with a comparison of dog, monkey, and man. *Psychol. Monogr.*, 1936, **47**, No. 212, 186–198.
12. HILGARD, E. R., & MARQUIS, D. G. Conditioning and learning. New York: Appleton-Century, 1940.
13. HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
14. LOUCKS, R. B. An experimental investigation of chained responses in indirect conditioning. *Psychol. Bull.*, 1939, **36**, 575–576.
15. LUMSDAINE, A. A. Conditioned eyelid responses as mediating generalized conditioned finger reactions. *Psychol. Bull.*, 1939, **36**, 650.
16. MILLER, J. The effect of facilitatory and inhibitory attitudes on eyelid conditioning. Ph.D. dissertation, Yale University. (Abstr. in *Psychol. Bull.*, 1939, **36**, 577–578.)
17. MILLER, J., & COLE, L. E. The influence of "voluntary" reaction upon the development and the extinction of the conditioned eyelid reaction. *J. genet. Psychol.*, 1936, **48**, 405–440.
18. PORTER, J. M., JR. The modification of conditioned eyelid responses by successive series of non-reinforced elicitations. *J. gen. Psychol.*, 1938, **19**, 307–323.
19. SEARS, R. R. Effect of optic lobe ablation on the visuo-motor behavior of goldfish. *J. comp. Psychol.*, 1934, **17**, 233–265.
20. SHIPLEY, W. C. An apparent transfer of conditioning. *J. gen. Psychol.*, 1933, **8**, 382–391.
21. SHIPLEY, W. C. Indirect conditioning. *J. gen. Psychol.*, 1935, **12**, 337–357.
22. WICKENS, D. D., & WICKENS, C. A study of conditioning in the neonate. *J. exp. Psychol.*, 1940, **26**, 94–102.
23. WICKENS, D. D., & WICKENS, C. Some factors related to the nature of pseudo-conditioning. *J. exp. Psychol.*, 1942, **31**, 518–526.

AN ATTEMPT TO APPRAISE INDIVIDUAL DIFFERENCES IN LEVEL OF MUSCULAR TENSION

BY M. A. WENGER*

Antioch College

Attempts have been in progress for many years to measure individual differences in skeletal muscular tension and to test the hypothesis that it bears a significant relationship to certain aspects of personality and to performance in specific situations. Recently Davis (**4**) has reviewed the literature concerning the methodology of measurement of tension and Courts (**2**) has discussed the evidence concerning its relationship to performance.

It is not surprising that neither Davis nor Courts offers a concise definition of muscular tension. As Davis points out, the concept has been used most frequently to designate a class of muscular processes, the interrelationships of which for the most part remain untested. In spite of the apparent dissimilarity among many of these so-called indices of tension, there is implied in the discussions pertaining to most of them an underlying common factor, which might be defined in terms of the number of discrete muscle fibers in a state of tonic contraction, or in terms of the amount of proprioceptive stimulation affecting the organism.

It is obvious that neither of these microscopic levels of description affords, at the moment, the opportunity for direct measurement. Of the six classes of indirect measurement of muscular tension described by Davis, the most direct form would be the recording of the electrical properties of muscles in terms of action potentials, the method to which Davis gives most attention. In the past, however, the majority of those experimenters who have employed this technique have contented themselves with the measurement of potentials from one or two muscle groups. Since differential relaxation of various muscle groups in the human organism is an accepted fact and since no one muscle group has yet been discovered which adequately reflects the level of tension in the majority of other muscles, it follows that no one placement of electrodes makes possible a reliable index of general muscular tension. It is possible that some critical muscle group may be found; or it is possible that a large number of electrode placements may yield a sufficiently homogeneous

* From the Samuel S. Fels Research Institute, Antioch College, Yellow Springs, Ohio.

body of data to obtain a reliable central tendency. In the meantime it is pertinent, perhaps, to approach the measurement of this general underlying muscular pattern in terms of another method of synthesis —factor analysis.

Certain reasons for expecting a relationship between general level of muscular tension and personality and a review of the pertinent literature were reported by Wenger in 1938 (**8**) along with the results of a preliminary experiment applying the method of factor analysis to the problem. A factor was found which gave evidence of muscular representation but the number of cases involved in the study was too small to permit generalizations. A later study of similar nature (**10**) involving 62 cases measured during the winter of 1940, also discovered a muscular factor which was comparable in certain respects to the one described earlier. Both studies suffered, however, in affording no clear definition of the factor. It was believed that if at least one test could be added to the battery which was unquestionably a test of generalized muscular tension it would serve to define and validate (or invalidate) the postulated factor of muscular tension.

Since no one test is available which all observers would agree constitutes a valid measure of general muscular tension, and since this concept at present is a subjective one with but very little objective support, it was decided to employ a rating scale. If it were found that observers could agree on a definition of muscular tension and then reliably rate behavior indicative of that trait there would be available a critical test for use in defining a muscular factor.

Two questions, then, were asked in the present study: (a) Can the postulated trait of characteristic level of general muscular tension be rated reliably? (b) If so, will the ratings serve to define a muscular factor when included in a body of appropriate data submitted to factor analysis? If both of these questions can be answered in the affirmative there is available at once a method for measuring individual differences in muscular tension in terms of factor scores.

First, a rating scale was developed after the form used by Champney (**1**) in which the trait to be rated is defined rather completely and cue statements are provided along a line 9 cm. in length. The rater is instructed to indicate his judgment of each subject by placing a check mark at the most appropriate point on the line. The resulting judgments may be scored in millimeters, as was done here, or any reduction thereof. A sample of the rating scale follows. Since the raters preferred to think in terms of muscular relaxation rather than in terms of tension, the scale is so named.

CHARACTERISTIC LEVEL OF MUSCULAR RELAXATION

Try to picture the child in non-emotional situations, and exclude the factor of restlessness as much as possible. If he is relaxed he appears to move freely; his arms and hands give the impression of just 'hanging'; his face is rather blank and immobile. If he is tense his movements are more jerky, more 'massive' in the sense of a 'tight' body reacting as an uncoordinated whole, or set ready to react at the slightest provocation. His face is not impassive. It and his whole body give an impression of strain.

Characteristic tension of the skeletal muscles is conceived of as a variable which may be separated from emotional behavior and from restlessness. Although emotion includes muscular tension, tension may be present in the absence of emotion. Although restlessness is based, no doubt, on a certain amount of residual tension, the tense child may inhibit restlessness while the relatively relaxed child may express more freely such overt reactions.

_____ Tense. Seems never to relax and take things easy. Looks and acts as though under constant muscular strain.
_____ Usually tense. Movements are often jerky, uneven.
_____ Median tension-relaxation.
_____ Relaxed, but not sleepy. Movements easy and 'loose'; rather limp.
_____ So relaxed he looks and acts almost asleep.

This rating scale was employed by eight staff members [1] of the Fels Research Institute for 74 children between 6 and 12 years of age. One of the raters (MP) felt incompetent to rate all of the children and therefore judged only the 53 subjects she knew. All others rated all of the cases and had been familiar with them over a period of at least 18 months. Seven of the raters were psychologists and one (MR) was a pediatrician.

Scores of individual raters ranged from 2 (tense) to 87 (relaxed) indicating that practically the entire scale was used by some of the raters. The mean ratings ranged from 17 to 75, with a combined mean of 39 and a σ of 12. Apparently these raters judged the majority of these children as characteristically more tense than relaxed, since the theoretical mean of the distribution would be 45. For only one of the raters was a higher mean demonstrated, a 46.

The first question to be considered is the comparability of the ratings. Did the raters agree in their judgments of muscular tension in these children to such an extent that the concept may be said to have meaning? To answer this question the mean of the scores by all raters have been correlated with the scores by each rater. The resulting coefficients are shown in the following tabulation.

Rater	HC♂	MW♂	TR♂	MR♂	MS♀	VN♀	FH♀	MP♀
r	.52	.64	.62	.51	.77	.45	.73	.75

It will be seen that in general the individual raters agree fairly well with the mean judgments. It is not surprising to find that rater HC, whose work is not directly with the children, shows a low coefficient. Raters MR and VN, on the other hand, had direct contact

[1] The author expresses his appreciation to those members of the staff who made the ratings. He also is indebted to the divisions of physical growth and psychology for the use of certain of their data.

with all the subjects but in rather specific situations: the first, during physical examinations, and the second during mental testing. The remaining five raters, who, to a greater extent, based their judgments upon more general observations, all show a considerable degree of correspondence. To approximate some measure of the actual range of reliability of the rating of this trait, the scores of the two raters correlating lowest with the mean (MR & VN) were intercorrelated, and likewise the scores of the two raters correlating highest with the mean (MS & FH).[2] The coefficients, when stepped up by the Spearman-Brown formula, were .76 in the first case and .92 in the second. The reliability coefficient of the mean ratings of eight raters who are familiar with the subjects rated may be expected therefore to fall between .76 and .92. It may be concluded that the trait of characteristic level of muscular tension can be rated with sufficient accuracy to make the ratings employed in this study constitute an acceptable criterion for the definition of a factor of muscular tension. It should be pointed out that this conclusion is valid only insofar as other observers, in addition to the eight raters of this study, agree that the trait as defined on this rating scale constitutes a valid description of general level of muscular tension.

The next step in the study was the factor analysis of data for 17 physiological variables obtained during the winter months of 1941 from the same 74 children mentioned above. The nature of the variables differed from those measured a year earlier (**10**), in four respects: some of the earlier tests were not repeated because of unreliability or impracticability; one of the blood pressure and one of the skin conductance measures were eliminated with the purpose of reducing the likelihood of specific factors appearing in the factor pattern; *Respiration Period* and *Sigma Respiration Period* were substituted for *Respiration Rate* and *Sigma Respiration Amplitude*; and five new variables were added. These new tests were *21 Vital Capacity*, as determined by the ordinary clinical spirometer, the best of three measures being recorded; *46 Body Index*, as determined by the formula weight/height, *69 Palm-Forehead Temperature Difference*, measured by means of the Hardy radiometer (**7**) from the center of the right palm and center of the forehead of the reclining subject at the end of approximately a half-hour experimental period; *71 Relative Lymphocytosis*, the percent of lymphocytes in the white cell blood count; and *72 Muscular Relaxation* (*rating*), the measure described in the first section of this paper. The methods employed for the other tests and the experimental procedure already have been reported (**10**) and are described in greater detail in a forthcoming report (**13**).

[2] Although rater MP showed the highest correspondence to the mean ratings, she judged only the 53 children best known to her.

TABLE I

THE INTERCORRELATIONS BETWEEN 17 PHYSIOLOGICAL MEASURES FOR 74 CHILDREN, 6–12 YEARS OF AGE, WITH AGE PARTIALED OUT

		21	22	23	25	28	29	33	37	39	44	46	57	69	71	72	77	80
21	Vital Capacity																	
22	Dermographia Latency	.09																
23	Dermographia Persistence	−.07	−.07															
25	Salivary Output	.09	−.01	.08														
28	Heart Period	.08	.16	.21	.18													
29	Sinus Arrhythmia	−.03	.10	.28	.08	.75												
33	Change in Palmar Log Conductance	−.22	.09	.08	−.02	.00	.04											
37	Standing Palmar Conductance	−.07	−.13	−.08	−.25	−.09	−.01	.03										
39	Non-Palmar Conductance	−.11	−.11	−.02	−.05	.03	.03	−.01	.05									
44	Sigma Respiration Period	.09	−.11	−.01	.02	.00	−.11	.16	−.03	.08								
46	Body Index	.04	−.38	.04	.05	.17	.04	−.21	−.03	−.15	−.01							
57	Respiration Period	.16	−.13	−.03	.01	−.05	−.11	.18	−.12	.07	.54	−.04						
69	Differential Skin Temperature	.07	.09	.02	−.04	−.09	−.10	−.06	.02	.37	.00	−.17	−.03					
71	Relative Lymphocytosis	.23	.02	.14	.27	.23	.13	−.09	−.06	−.01	.06	−.09	.07	.15				
72	Muscular Relaxation (rating)	−.12	.02	.08	.11	.08	.09	.07	.10	−.08	.04	.39	.34	−.08	.15			
77	Diastolic Blood Pressure	.00	−.10	.14	−.11	−.07	.08	−.04	−.06	.05	.08	.27	−.06	.13	.23	.13		
80	Pulse Pressure	.04	.09	−.19	.16	.02	.15	.11	−.07	−.05	−.10	−.22	.10	−.07	−.14	−.01	−.76	
	Age	.78	−.16	.13	.27	.41	.17	.40	−.04	−.30	−.05	−.06	.10	.05	−.12	.18	.39	−.08

The intercorrelations of the results of the 17 tests, with chronological age partialed out, are shown in Table I and the centroid pattern of the factor analysis may be seen in Table II. Although there was some indication of significant third and fourth factors, attempts to rotate four factors produced ambiguous results. As was the case in the analysis of the 1940 data (**10**), however, the first two factors appeared to yield a meaningful pattern. Factor L_c may be regarded as arising largely because of the high intercorrelation between *77 Diastolic Blood Pressure* and its partial derivative, *80 Pulse Pressure*. Factor L_d has a wider representation, however, and it might be argued that it should have been included in the rotations. On the other

TABLE II

THE CENTROID PATTERN OF THE FACTOR ANALYSIS

Variables		Factors			
No.	Name	L_a	L_b	L_c	L_d
21	Vital Capacity	.12	.16	.10	−.20
22	Dermographia Latency	.03	.34	−.26	−.15
23	Dermographia Persistance	.23	.04	.25	.04
25	Salivary Output	.35	.23	.05	−.11
28	Heart Period	.61	.36	.24	.24
29	Sinus Arrhythmia	.53	.32	.18	.35
33	Change in Palmar Log Conductance	.10	−.14	−.33	−.06
37x	Low Standing Palmar Conductance	.27	.15	.03	−.24
39x	Low Volar Conductance	.25	.00	−.20	.16
44	Sigma Respiration Period	.26	−.42	−.11	−.38
46	Body Index	.18	−.31	.33	.42
57	Respiration Period	.36	−.44	−.31	−.33
69x	Small Differential Skin Temperature	.21	−.10	−.23	.37
71	Relative Lymphocytosis	.32	.18	.27	−.31
72	Muscular Relaxation (rating)	.42	−.37	−.06	.18
77	Diastolic Blood Pressure	.10	−.39	.66	−.23
80	Pulse Pressure	.04	.40	−.63	.17

Symbol x signifies reflection of trait.

hand, in view of the low magnitude of the coefficients of Table I it seems improbable that more than two factors could be described reliably. All four factors have been reproduced, however, should the present two-factor solution be questioned.

The rotated solution, with the supposed muscular factor Mu from the 1940 analysis are shown in Table III. Factor L_1 will not be considered here since it has been discussed elsewhere (**12**). It was found to be similar to the autonomic factor *Nu* from the 1940 analysis, and more recently both factors have been shown to correspond closely with a factor found for data collected during the winter of 1942 (**14**). It is considered that the existence of a factor representing the functional state of the autonomic nervous system has been demonstrated for this group of children.

Factor L_2 is seen to have a high correlation with *72 Muscular Relaxation* (rating), and to be related significantly also to *57 Respiration Period* and *44 Sigma Respiration Period.* This picture of the person rated as tense also manifesting a rapid and regular pattern of respiration is not surprising and affords additional justification for calling factor L_2 a muscular factor. Although the factor is seen to be similar in certain respects to factor Mu from the earlier analysis, some marked differences do appear.

The greatest difference is that manifested by variable *33 Change in Palmar Log Conductance.* This measure deserves some explanation. Following the findings of White (**15**) Wenger and Irwin (**9**)

TABLE III

THE TWO-FACTOR PATTERN FOR THE 1941 DATA WITH FACTOR MU (REFLECTED) FROM THE 1940 ANALYSIS

	Physiological Measures in School-Age Children	L_1	L_2	Mu_3
21	Vital Capacity	.19	.07	
22	Dermographia Latency	.21	.27	.31
23	Dermographia Persistance	.22	−.09	−.28
25	Salivary Output	.42	.01	.05
28	Heart Period	.71	−.02	.62
29	Sinus Arrhythmia	.62	−.01	.51
33x	Small Change in Palmar Log Conductance	.03	.17	.74
37x	Low Standing Palmar Conductance	.31	−.02	−.03
39x	Low Volar Conductance	.21	−.13	−.08
44	Sigma Respiration Period	.00	−.49	.19*
46	Body Index	−.01	−.36	
57	Respiration Period	.07	−.56	.30
69x	Small Differential Skin Temperature	.13	−.20	
71	Relative Lymphocytosis	.37	−.02	
72	Muscular Relaxation (rating)	.16	−.54	
77x	Low Diastolic Blood Pressure	.12	.38	.50
80	Pulse Pressure	.25	.32	.23

* σ Respiration Amplitude.

and Freeman and Simpson (**6**), that the electrical resistance of the palms decreases during induced tension in the skeletal musculature and increases during muscular relaxation, the amount of change in resistance between the mean of the standing measures and the mean of the reclining measures, equated from subject to subject by conversion to log conductance change (**3**), has been regarded as a measure of the amount of muscular relaxation during the experimental period. At the time of the 1940 analysis this measure seemed the clearest one available in terms of which to define factor Mu. For the present analysis, however, *72 Muscular Tension* (rating) was allowed to determine the axis of the factor. If Factor L_2 is a valid description of a factor representing characteristic level of muscular tension and it correlates only .17 with the temporal sample represented by *33*

Change in Palmar Log Conductance, it is apparent that the latter is at least an unreliable indicator of general muscular tension, and that factors L_2 and Mu could not be expected to be identical. There can be little question but that Factor L_2 will prove to be the more significant of the two factors. It is of incidental interest that their similarity and difference appear to arise from the fact that variable *74* is essentially a measure of status, while test *33* is a measure of change. The child who is either tense *or* relaxed throughout the experimental period does not show change and has a low score on test *33*. Only the child who shifts from tension to relaxation shows a high score. Since high scores on factor Mu would never include the child who is always relaxed, but would always include children who manifest some tension, the basis for the lack of correspondence between the factors is apparent. Actually, except for the criterion measures themselves, only one test serves to differentiate the two factors, *57 Respiration Period*. It will be seen (Table III) that the child rated as relaxed tends to demonstrate a slow respiration rate whereas the child scoring high on Variable *33* would tend to have a fast respiration rate.

Before describing the attempt to measure individual differences for Factor L_2 it is of interest to consider the correlations and factorial representation of the other new tests in the battery. Variable *21 Vital Capacity* is seen to have a low relationship to both factors and to all variables except chronological age. No. *46 Body Index* bears no relationship to the autonomic factor but is related to Factor L_2. Apparently the child underweight for height is more apt to be rated as tense than relaxed, and to manifest the remainder of the pattern described by Factor L_2. Looking at Table II, *Body Index* is seen to show some correlation with the two unrotated factors, L_c and L_d, a fact which might be regarded as reason for including all four factors in the rotated solution.

It is surprising that *69 Differential Skin Temperature* does not have a higher representation on the autonomic factor, for it is frequently regarded as a test of autonomic function. It is seen, however, to bear a low relationship to the other tests, and to find its highest correlate in *21 Vital Capacity*. Measure *71 Relative Lymphocytosis* was included because of a finding by Farris (**5**) that percent lymphocytes increases after certain emotional reactions. It was believed possible that an increase over normal percent for age might be found in individuals manifesting a functional predominance of the sympathetic nervous system. The finding is contrary to expectation. Instead, a high relative lymphocytosis appears to be associated with a pattern of parasympathetic predominance, and to be unrelated to other than the autonomic variables. The intercorrelations of the

remaining tests in the battery may be compared by the reader with those for the 1940 data (10).

The next step in the study was to measure what are presumed to be individual differences in muscular tension as indicated by differences in factorial estimates for L_2. The method employed was the same as that already utilized in the measurement of individual differences for the autonomic factor (11). By means of the Doolittle method of multiple correlation a regression equation was obtained which weights the tests according to their factorial representation. Instead of using all of the tests in the battery, the method of partial estimation was employed, involving only the tests which correlate highest with the factor. In this instance all tests were used which correlated higher than .25 with Factor L_2, but for the following exceptions. Although both blood pressure variables satisfy this criterion they also manifest a high intercorrelation so, rather than unduly weight the scores in this direction, only one of them was utilized. Since *33 Change in Log Conductance* is regarded as being a valid measure of a least temporal muscular tension, it was included. Variable *46 Body Index* is a more static variable than the others and its inclusion would tend to make sequential measurements stable and reduce the possibility of disclosing real changes in behavior with changes in the environment. Since the analysis of environmental influence is one of the primary purposes of the Fels study, this variable was eliminated from further consideration.

Six tests were involved, then, in the multiple correlation: *22 Dermographia Latency*, *33 Change in Palmar Log Conductance*, *44 Sigma Respiration Period*, *57 Respiration Period*, *72 Muscular Relaxation* (*rating*) and *77 Diastolic Blood Pressure*. It will be noted in the resulting regression equation which follows that the five variables showing negative correlations with Factor L_2 have been reflected so that all signs are positive. An *x* in subscript indicates this reflection of the trait.

$$\bar{L}_2 = .19z_{22} + .06z_{33x} + .28z_{44x} + .25z_{57x} + .40z_{72x} + .30z_{70x}$$

The equation is to be read: The estimated factor score for a given individual equals .19 times his standard score for *22 Dermographia Latency* plus .06 times his reflected standard score for *33 Change in Palmar Log Conductance*, etc. High scores from this equation should denote a high degree of muscular tension while low scores should indicate those individuals who characteristically are muscularly relaxed.

This equation was then applied to the data for the 74 children involved in the study. The resulting distribution of factor estimates is shown in Fig. 1. It will be seen that although the distribu-

tion is broad, there is a definite skew with the greatest concentration of cases on the 'tension' side of the mean. Apparently the factor estimates agree with the ratings in describing this group of children as more tense than relaxed. Fig. 1 shows also the correspondence between the two measures in the shape of distribution. Both sets of scores are presented in standard score form in order to compare them on the same scale.

The question now arises, which is the better measure of muscular tension? It is obvious that the problem can only be approached

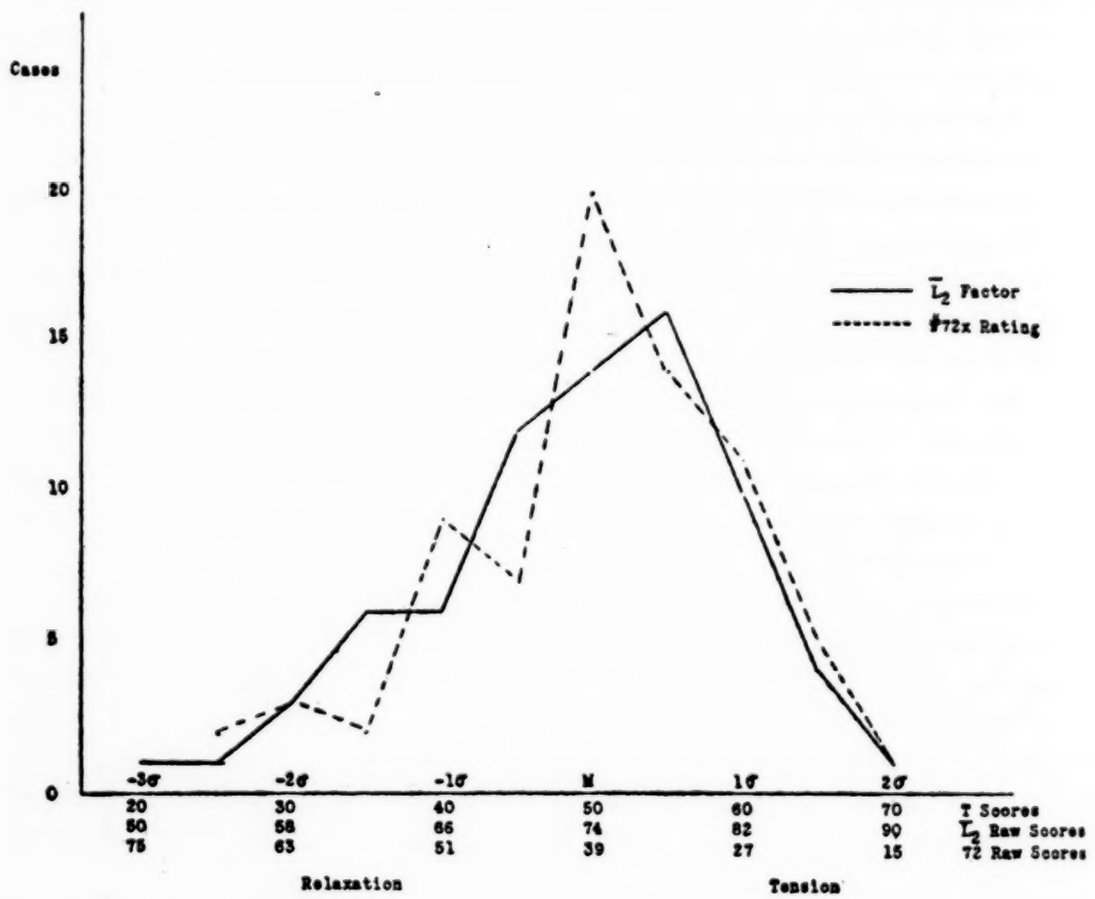


FIG 1. Frequency distributions for two measures of muscular tension for 74 children 6–12 years of age.

indirectly, since there is no outside criterion. One is valid only insofar as observers agree on the trait definition, and then suffers by the difficulty of the rating technique itself. The other, while affording a certain internal consistency which may be interpreted as a manifestation of validity, is defined chiefly in terms of its correlation with the former. The comparative validity of the two measures then can only be determined by the degree of relationship to other traits or factors which are known or postulated to be related to muscular tension. Since for this type of description of muscular

tension no correlates are yet known, the comparative validity of the present measures must rest on their relationship to postulated correlates.

These postulated correlates of tension are chiefly of two general kinds: changes in efficiency of performance; and changes in various aspects of personality. The former have been based largely on the results of experiments dealing with induced tension and for the most part have not involved the concept of characteristic level of tension. The latter have evolved around just this concept, however, so it is the area of personality that we have turned here. Ratings of certain personality traits made during the summer of 1940 were available for 44 of the children of this study, and from them were selected the traits which, according to the arguments previously advanced (**8**),

TABLE IV

THE RELATIONSHIP OF NINE PERSONALITY RATINGS TO TWO MEASURES OF MUSCULAR TENSION FOR 44 CHILDREN

	Personality Traits	Correlations with	
		$\bar{L}_2$	r_{2x}
III	Energeticness	.46	.46
IV	Fatigueability	−.35	−.24
V	Frequency of Emotion	.46	.47
VI	Restlessness	.48	.50
VII	Sensitiveness	−.10	−.02
IX	Emotional Control	−.33	−.32
X	Distractability	.19	.17
XI	Impulsiveness	.35	.41
XII	Carelessness	.22	.21

might be expected to correlate with these measures of muscular tension. The trait names and their correlations with each of the two scores of muscle tension are shown in Table IV.

It will be seen that in general the estimates for Factor L_2 show a degree of relationship to these personality variables comparable to that shown by the ratings of muscular tension alone. The correlations with $\bar{L}_2$ may be regarded then as furnishing some evidence of the validity of $\bar{L}_2$ as a measure of characteristic level of muscular tension in addition to that already furnished by the relationship between the factor and the ratings; and furthermore, some additional support is furnished to the hypothesis that muscular tension is related to certain aspects of personal-social behavior.

It may be concluded, since $\bar{L}_2$ is a more objective measure than are the ratings alone, that $\bar{L}_2$ furnishes the best available method for measuring individual differences in characteristic level of muscular

tension. Future factor analyses of tests more rigorously designed to measure the many ramifications of muscular tension should discover a similar but more comprehensive muscular factor of considerable significance in the appraisal of human behavior.

Summary

Two questions were asked in this study: (a) can the postulated trait of characteristic level of muscular tension be rated reliably; and (b) if so, will the ratings serve to define a muscular factor when included in a body of appropriate data submitted to factor analysis?

A rating scale was devised and used by 8 raters in appraising this characteristic for 74 children 6–12 years of age. It was found that the data of individual raters correlated from .45 to .75 with the means of the ratings and that the reliability of mean judgments of the eight raters was between .76 and .92. It was concluded that characteristic level of muscular tension could be rated sufficiently reliably for the purposes of this study.

The ratings and 16 other physiological variables were then intercorrelated and submitted to factor analysis by the centroid method. At least two meaningful factors were found, one of which was defined as a factor of muscular tension. Factorial estimates of this factor for the 74 cases yielded a distribution of scores markedly similar to that found for the ratings alone, and were found to correlate equally well as the muscular ratings with ratings of certain personality traits whose relationship to general muscular tension had been postulated. Both sets of correlations are believed to furnish additional evidence that generalized muscular tension is a correlate of certain aspects of personality.

It was concluded that factorial estimation affords the most adequate technique for the measurement of general muscular tension, and that further analyses of more rigorously designed tests should yield a muscular factor of considerable significance in the appraisal of human behavior.

(Manuscript received October 15, 1942)

References

1. Champney, H. The measurement of parent behavior. *Child Dev.*, 1941, **12**, 131–166.
2. Courts, F. A. Relations between muscular tension and performance. Psychol. Bull., 1942, **39**, 347–367.
3. Darrow, C. W. The equation of the galvanic skin reflex curve: I. The dynamics of reaction in relation to excitation background. *J. gen. Psychol.*, 1937, **16**, 285–309.
4. Davis, R. C. Methods of measuring muscular tension. *Psychol. Bull.*, 1942, **39**, 329–346.
5. Farris, E. J. Increase in lymphocytes in healthy persons under certain emotional states. *Amer. J. Anat.*, 1938, **63**, 297–323.

6. FREEMAN, G. L., & SIMPSON, R. M. The effect of experimentally induced muscular tension upon palmar skin resistance. *J. gen. Psychol.*, 1938, **18**, 319–326.
7. HARDY, J. D. The radiation of heat from the human body: I. An instrument for measuring the radiation and surface temperature of the skin. *J. clin. Invest.*, 1934, **13**, 593–604.
8. WENGER, M. A. Some relationships between muscular processes and personality and their factorial analysis. *Child Developm.*, 1938, **9**, 261–276.
9. WENGER, M. A., & IRWIN, O. C. Fluctuations in skin resistance of infants and adults and their relation to muscular processes. In Studies in Infant Behavior, III. *Univ. Iowa Stud. Child Welf.*, 1936, **12**, No. 1. Pp. 207 (pp. 7–90).
10. WENGER, M. A. A study of physiological factors: the autonomic nervous system and the skeletal musculature. Human Biol., 1942, **14**, 69–84.
11. WENGER, M. A. The measurement of individual differences in autonomic balance. *Psychosom. Med.*, 1941, **3**, 427–434.
12. WENGER, M. A. The stability of measurement of autonomic balance. *Psychosom. Med.*, 1942, **4**, 94–95.
13. WENGER, M. A., & ELLINGTON, M. The measurement of autonomic balance in children: method and normative data. (To be published.)
14. WENGER, M. A. A further note on the measurement of autonomic balance. (To be published.)
15. WHITE, M. M. Relation of bodily tension to electrical resistance. *J. exp. Psychol.*, 1930, **13**, 267–277.

DEVELOPMENT OF DIFFERENTIAL APPETITE IN THE RAT *

BY LEON FESTINGER

State University of Iowa †

That organisms prefer one thing to another is a common observation. We have little knowledge, however, of the psychological dynamics behind such preferential behavior. This experiment is an attempt to study one factor which might bring about such 'differential taste'; namely, differences in previous experience. More specifically, the question is: if an animal is given more of one food than of another, will it eventually come to prefer that food on which it has experienced relative deprivation?

Procedure

Rats were used as subjects so that feeding habits and general hunger drive could be controlled.

The apparatus used is shown in Fig. 1. 'S' is the starting box; movable doors are indicated by 'd.' Three groups of animals were run in this apparatus under the following conditions:

1. Experimental Group: (N = 10).

The two experimental foods chosen were powdered whole wheat grain and dry milk.[1] The former was placed in the food box of Alley A, the latter in the food box of Alley B.

The subjects were hooded rats of the Western Reserve 'non-emotional' strain. The animals, under 12 hours' deprivation of food, were given 4 runs per day for 24 consecutive days. Every fourth day they were allowed to choose; that is, doors to both alleys were open. On all the other days, the runs were forced (only one door would be open) so that the animal ran twice to each alley each day. These forced runs were given in an ABBA order on one day, BAAB order the next, and so on. Thus in the 24 days, the animals had 6 days of free runs to food. Five of the rats were allowed to feed for one minute upon going to Alley A, but only for 10 seconds when they ran to Alley B. The other 5 rats were given the one-minute feeding in Alley B, and the ten-second feeding in Alley A.

After the 24 days described above, each animal was given free runs in the apparatus under 0, 6, 18, 24, and 36 hours of deprivation of food.

2. Control Group A: (N = 6).

These rats were Albinos from the stock of the Department of Psychology of the University of Iowa. They were run in the same manner as the experimental group except for one difference. For this group, the powdered whole wheat grain was placed in both alleys. The purpose of this group was to establish whether rats could discriminate between ten-second and one-minute feeding.

3. Control Group H: (N = 9).

---

* The author wishes to acknowledge his indebtedness to Dr. K. W. Spence, Head, Department of Psychology, State University of Iowa, for his aid and suggestions, and to Mrs. Nancy Phillips Perkins for gathering the data for the control group H.

† From the Iowa Child Welfare Research Station.

[1] P. T. Young (5) found that rats in a free-choice situation liked these two foods equally well, that is, ate them both equally often.

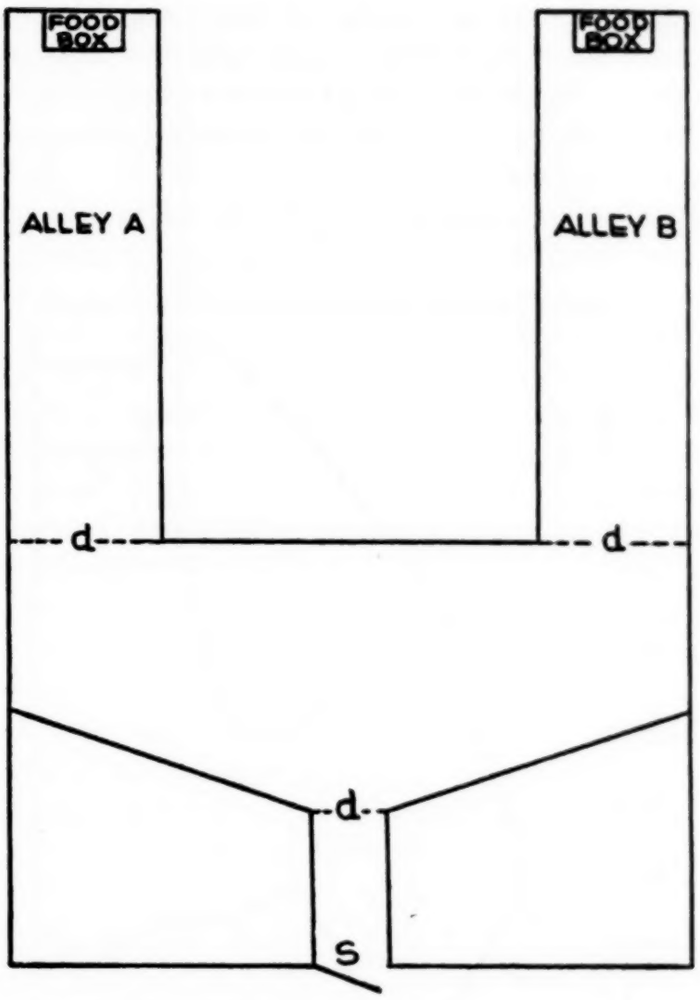


FIG. 1. Experimental apparatus.

These rats were bred from the animals in the experimental group and were treated exactly as control group A. The purpose here was to be able to compare rats of the same stock.

All groups were fed regularly in their cages on the usual laboratory food employed. The experimental foods were encountered only in the experimental apparatus.

RESULTS AND DISCUSSION

The results are presented in terms of the mean number of runs made to the food on which one minute of feeding was allowed. Since 4 runs were made on each day, a mean of 2.00 represents running equally often to both alleys, and a mean of 4.00 represents running every time to the 'one-minute food.'

The results are shown in Fig. 2. The experimental group, on the first day of free runs, has an average of 2.6 runs to the 'one-minute food.' (It should be remembered that the first day of free runs was the fourth day of running in the apparatus.) On the second and third day of free runs, the runs to the 'one-minute food' drop. They rise again on the fourth day, and on the fifth and sixth days of free runs,

both show an average of 2.5 runs to the 'one-minute food.' On both the fifth and sixth days there was only one subject that ran all 4 times to one side. The low number of runs to the 'one-minute food' is not, then, the result of some animals running always one way and others running always the other way.

One would ordinarily expect that if the animal wants food, and if the animal has learned where food is to be obtained, then it will

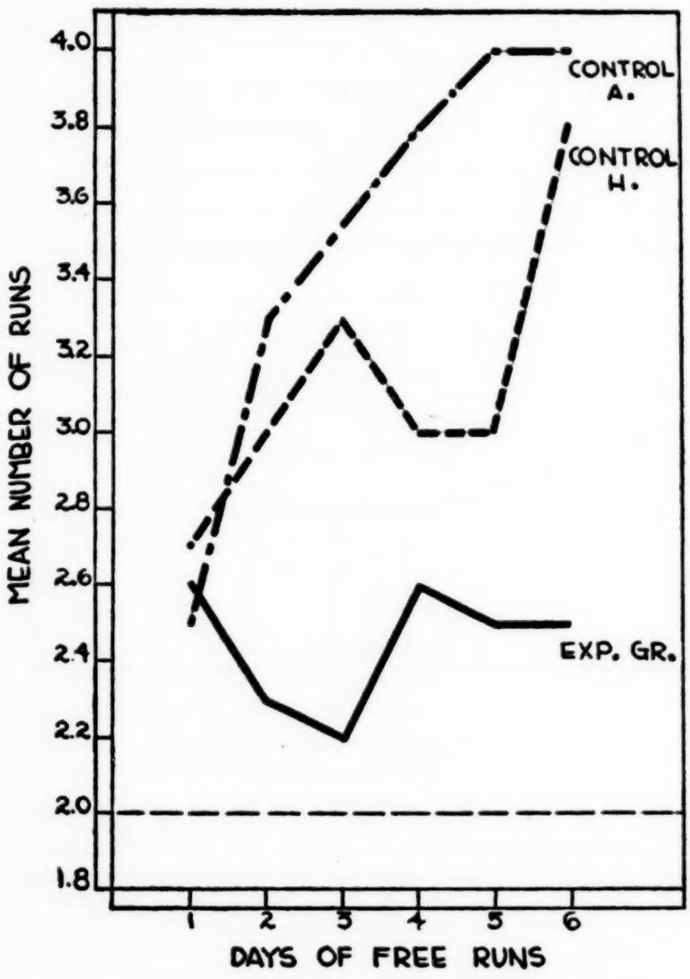


FIG. 2. Mean number of runs to 'One-Minute Food' on days of free runs.

run to the place where it can obtain a greater quantity. The experimental group does not do this. Is it because the tendency to go to a greater amount of food is counterbalanced by a tendency to go to 'less food' because it has become more desirable? Or is it simply that the animals have not learned to discriminate between one minute and 10 seconds of feeding?

The question of whether rats can discriminate between one-minute and ten-second feeding can be answered directly by means of the control groups.

From Fig. 2 it can immediately be seen that both control groups

learn to run to the 'one-minute food.' On the first day of free runs, the differences between the experimental and the two control groups are not significant. The *t*-values for differences between the means of the experimental group and control A and control H are .23 and .20, respectively. From then on, however, the control groups run consistently more often to the 'more food' than the experimental group.

The statistical significance of this difference between the experimental group and the control groups can easily be demonstrated with the use of the binomial expansion. If there is really no difference between the experimental group and the control groups, then the chances of having the control groups superior on five consecutive test days are one out of thirty-two. The difference between the experimental group and either of the control groups is significant at the 3.1 percent level of confidence.

The control groups demonstrate two interesting things. In addition to showing that rats can discriminate between one-minute feeding and ten-second feeding, they also show that when presented with a choice between a greater and a lesser amount of food, the animal will run to the greater amount. The effect of the amount of food offered upon the behavior of the organism has been demonstrated by Grindley (**3**), Gagne (**2**), and Fletcher (**1**). In all of these experiments, however, no choice was involved, and the measure taken was the time of running to the food, which in all experiments was found to decrease as the amount of food offered was increased. The data of the two control groups in this experiment demonstrate that the effect of offering a greater amount of food also holds for choice or discrimination situations.

If we then accept the fact that the rat will run to the greater amount of food when given a choice, and if we also accept the fact that the rats are able to discriminate between the one-minute and the ten-second feeding in our experimental situation, we must come to the following conclusion. Some factor, acting opposite to the tendency to run to the greater amount, is operating with respect to the experimental group so as to produce the observed result (to reduce the number of runs to the greater amount of food).

The only difference between the experimental and control groups is the fact that the control groups had the same food in both alleys and the experimental group had different foods in the two alleys. For the experimental group, there exists a food of which they have been relatively deprived. If this food has become more desirable relative to the other experimental food, then the results obtained should be expected. The increased desirability of the 'less food' would act counter to the tendency to run to the greater amount.

If this interpretation is correct, the following should be correct. The more intense the general hunger drive becomes, the smaller role should the differential appetite play in the choice; that is, quantity should become a more important factor. The less intense the general hunger drive becomes, the more important should the differential taste become. If the explanation offered is correct, we may then make the prediction that for the experimental group the number of runs to the 'one-minute food' should increase as the intensity of the hunger drive increases.

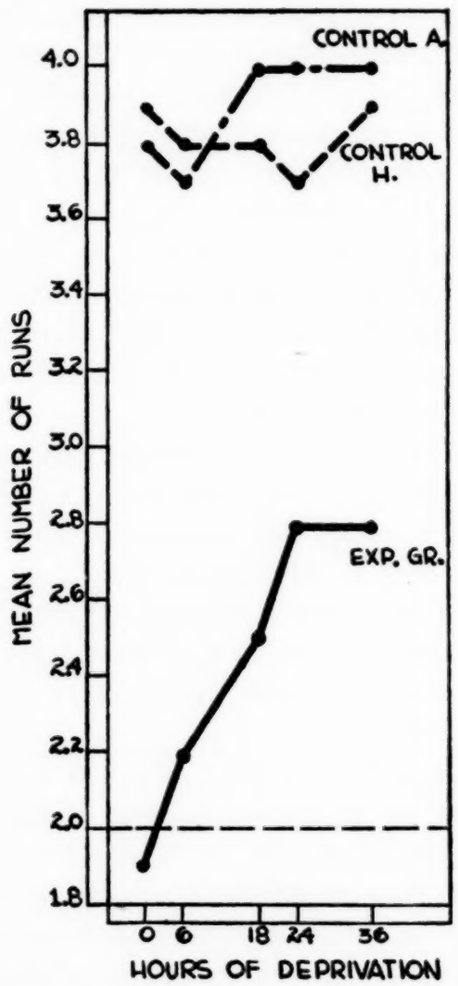


FIG. 3. Mean number of runs to 'One-Minute Food' under different hours of deprivation of laboratory food.

Fig. 3 presents the data for the runs under different numbers of hours of deprivation of food and shows this prediction to be correct. Under zero hours of deprivation of food, the animals run 1.9 times to the 'more food' on the average. This is less than they have ever run before. The number of runs to 'more food' increases steadily with increase in the number of hours of deprivation until it reaches a value of 2.8 at both 24 and 36 hours of deprivation. This last is more than these animals had ever run in the past.

The control groups, on the other hand, maintain essentially the same level of performance under different hours of deprivation. The difference between the experimental and control groups is even more striking here than before. We are then left with the plausible interpretation that during a period of relative deprivation one food has become more desirable than another, whereas previously they were both equally desirable.

Not directly pertinent to the main question involved, but worthy of mention, is the question of whether the 'habit pattern' acquired by the control group was exhibited under zero motivation. It can hardly be said that zero hours of deprivation indicates zero motivation. The feeding occurred in the cages on the usual laboratory food and it cannot be said that the need for the experimental food was satisfied completely by the other food.

In fact, the whole experiment tends to show that needs for food can exist other than that of general hunger need. Corroborative of this point of view is the fact that all the animals, experimental and control groups both, ate the experimental foods under zero hours of deprivation of the laboratory food. The decrease in motivation, which undoubtedly was present, is shown by the fact that all the rats ran more slowly under zero hours of deprivation than under other conditions.

Theoretical Interpretation

Although the results can be made meaningful in common-sense terms, a formal theoretical explanation presents certain difficulties. There are two major findings which are in need of explanation. The first is the fact that the rats did not run always to the 'more food' when a different food in lesser quantity was available. The second fact is that increasing the general hunger drive increased the number of times the rats did run to the 'more food.'

In common-sense terms, we may say plausibly enough that the animals acquired a relatively greater liking for one food because they got less of it, and we may make it more plausible by drawing analogies to people who like things which they are unable to get. This, however, offers no clue to the dynamic relationships behind the development of this differential preference. With respect to the second result, we may say that the more intense the general need for food becomes, the greater role quantity should play, and the less influential taste becomes in determining a choice; therefore, the rats would choose the greater quantity more often when the hunger need gets stronger. This again amounts merely to a restatement of the observed fact.

It seems desirable to be able to explain the obtained results in

terms of the constructs of some theoretical system. Although this may present difficulties, an attempt in this direction should prove of value. It may at least be possible to indicate the direction which such theories must take in the formulation of their empirical laws to handle data of this nature.

Some constructs from the theoretical system of Lewin (4) seem to be of such nature as to be able to handle the results of this experiment. The principal reference is made here to the construct of tension system. The individual possesses various needs. We may speak of these needs as representing separate regions in the person. These regions may be more or less differentiated or subdivided. As an example, we may point to the hunger need which may be subdivided into a need for meat, a need for vegetables, etc. A system such as this, when unsatisfied, is said to be in tension. The construct of tension has another property, namely that of spreading. As time progresses, the tension in a system tends to spread in such a way as to become equal to that of the surrounding regions. One of the factors influencing the rate of spread is the strength of the boundaries separating the regions involved. The weaker the boundary, the more rapid will be the spread of tension. The question of strength of boundary can, in many cases, be looked at as a matter of discrimination. The more clear the discrimination, the more clear will be the separation between the regions, or the stronger the boundary between them.

When a system is in tension, a given realm of objects or activities becomes attractive in consequence of the tension. This can be illustrated by another example of hunger. When a person is not hungry (system not in tension), food may not seem particularly attractive to him. As the person becomes more and more hungry (system in tension), not only do foods become more attractive to him, but a greater number of foods become attractive. In a more formal manner, we may say that coordinate to the existence of a system in tension is the existence of regions of positive valence which will satisfy this tension, and furthermore, the magnitude of valence is a monotonic function of strength of the tension.

Our experimental situation and procedure can be stated in terms of the constructs outlined above. Because of the experience with the two experimental foods and as a consequence of discriminating between them, the general hunger drive becomes differentiated so that specific tension systems exist for each of the two experimental foods. Since the animals' appetites for these two foods are equal at the start of the experiment, we may say that these two tensions, if differentiated, would be equal to begin with.

It is probably necessary to state here that the needs for the two

experimental foods have two components. One of these is due to hunger (a nutrient need); the other must be regarded as a non-nutrient need (due to taste and many other factors). At this point we are confronted with the problem of the interaction between these two components. We shall, for the purpose of the present discussion, make the simple assumption that the fluctuations in general hunger add or subtract equally from the total tension in all sub-systems.

When one states that the amount of tension reduction in these sub-systems is proportional to the amount of the corresponding food consumed, one can immediately see the effect of the experimental procedure. Starting with the two systems equal in tension, we continually reduce one by a greater amount than the other. The tension reduction in the sub-system corresponding to the 'one-minute food' is greater than the tension reduction in the sub-system corresponding to the 'ten-second food.' As a result, at the end of one day's runs, the sub-system corresponding to the 'ten-second food' is in a higher state of tension than the 'one-minute food' sub-system. After repeated trials, the difference in tension between the two sub-systems will fluctuate within a stable range, depending upon the strength of the boundary separating the two regions. Of course, if the animal has not discriminated between the two foods, the boundary is non-existent, and there will be no difference in tension.

Co-existent with the difference in tension, a *unit* of that food corresponding to the sub-system with the greater tension will be more attractive (have a greater positive valence) for the animal. Since the less attractive food exists in greater quantity, there will be a conflict of forces resulting in fewer runs to the 'one-minute food' than would be the case if both were equally attractive.

This analysis may also be used to derive the increase of runs to the greater amount of food with increase in the general hunger need. If an increase in the general hunger need increases the tension in both of the sub-systems under consideration, then we may regard it as increasing the forces acting on the boundaries of the sub-regions. Thus, by increasing the hunger need, we increase the probability that these forces acting on the boundary will be stronger than the maximum boundary forces. If such a situation obtains, then the general hunger need obscures the discrimination and the animal should run to the greater quantity of food. Therefore, as we increase the hunger drive we should, as we do, observe an increase in the number of runs to more food.

An explanation of the obtained results is also feasible in S-R terms.[2] Through the training process we eventually come to a state

[2] This explanation was suggested by Dr. R. R. Sears, Director, Iowa Child Welfare Research Station, State University of Iowa. In what follows, S_1 will refer to the stimuli of turning at

of affairs such that S_1, S_D, and S_{rg_1} are associated with R_1 (eating the 'one-minute food') while S_2, S_D, and S_{rg_2} were associated with R_2 (eating the 'ten-second food'). The habit strength is stronger in the former due to the greater amount of reinforcement. One must here assume that the relative deprivation or frustration of R_2 makes the anticipatory goal responses for the 'ten-second food' stronger than in the case of the 'one-minute food.' Therefore, because of the stronger stimulus complex, the animal will run less often to the 'one-minute food' than if S_{rg_1} and S_{rg_2} were equal in strength.

The increase in the number of runs to the 'one-minute food' with increasing hunger drive is also taken care of, since that means an increase in S_D which is common to both R_1 and R_2, therefore making the total difference in response tendency more in favor of the 'one-minute food.'

Summary and Conclusions

A group of rats were run in a single discrimination point apparatus. The choice involved was between a ten-second feeding on one food and a one-minute feeding on a different food. The appetites for the two foods used were equal at the start of the experiment. After 24 days, the animals ran only slightly more than 50 percent of the time to the greater amount of food. A control group showed that if only one food were involved, the animals would learn to run to the greater amount of food almost 100 percent of the time.

It was further found that the number of runs made by the experimental group to the greater amount of food was increased by increasing the strength of the hunger drive.

An interpretation was made in terms of the one food's becoming more attractive to the animal because of the relative deprivation.

(Manuscript received October 19, 1942)

References

1. Fletcher, F. M. Effects of quantitative variation of food incentive on the performance of physical work by chimpanzees. *Comp. Psychol. Monogr.*, 1940, **16**, 3, pp. 46.
2. Gagne, R. M. The effect of spacing of trials on the acquisition and extinction of a conditioned operant response. *J. exp. Psychol.*, 1941, **29**, 201–216.
3. Grindley, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, **20**, 173–180.
4. Lewin, K. *Principles of topological psychology.* New York: McGraw-Hill, 1936. Pp. 231.
5. Young, P. T. Preferences and demands of the white rat for food. *J. comp. Psychol.*, 1938, **26**, 545–588.

the choice point in the direction of the 'one-minute food'; S_2 will refer to similar stimuli in the direction of the 'ten-second food'; S_D will refer to the hunger drive stimuli; R_1 and R_2 will refer to the responses of eating the one-minute food and eating the ten-second food respectively; S_{rg_1} and S_{rg_2} (one-minute food and ten-second food, respectively) will refer to the stimuli aroused by the anticipatory goal responses made by the subject before the goal is reached.

STUDIES IN SERIAL VERBAL DISCRIMINATION LEARNING: III. THE INFLUENCE OF DIFFICULTY ON REMINISCENCE IN RESPONSES TO RIGHT AND WRONG WORDS [1]

BY DAVID C. McCLELLAND

Wesleyan University

Previous experiments in this series (**5**, **6**) have shown that reminiscence appears in discrimination learning and that it apparently is localized in the wrong stimulus members of the pairs of words presented. The method of testing for reminiscence in the experiment which determined its locus (**6**) was to combine the wrong words from half the original pairs and the right words from the other half with new words and to use these new combinations for relearning. Since it was easier after a 2-minute rest than after a 6-second rest to learn the pairs to which the wrong words had been transferred, and not easier to learn the pairs to which the right words had been transferred, it was concluded that reminiscence was confined to the transferred wrong words.

These results were interpreted as supporting the hypothesis that reminiscence is due to recovery from performance inhibition rather than to other factors such as the differential rate of forgetting of right and wrong responses (**1**, **8**) or maturation (**9**). They did not support such a conclusion with absolute certainty, however. The argument was weakened by the fact that at the beginning of relearning the right word pairs were so much easier to master than the wrong word pairs. Such a difference in difficulty, while it was the basis of the conclusion that it was released from inhibition which was involved, could also have explained why the other theories were not confirmed. That is, the right word pairs were so well mastered at the time of transfer that they may not have been able to show the forgetting required by the differential rate theory or the growth demanded by the maturation hypothesis.

It was to meet this weakness in the case for a performance theory that the present experiment was performed. Its purpose was to provide right and wrong parts of a discrimination which would be

[1] This paper is a modification of part of a dissertation presented in June, 1941 to the Faculty of the Graduate School of Yale University in partial fulfillment of the requirements of the degree of Doctor of Philosophy. I am greatly indebted to Dr. C. I. Hovland for his many helpful suggestions in the planning, execution, and interpretation of the research.

equally hard to relearn. To accomplish this purpose the transfer method had to be discarded. Instead, the subjects were presented with a list of words one at a time rather than in pairs. They had to learn to say some of them (the right words) and not to say others (the wrong words). With the discriminations divided in this way so that a separate response was required to each part, retention of the right and wrong parts could be tested separately, undisturbed by the factor of difference in degree of learning at the time of testing.

Procedure

Two lists were constructed of 20 two-syllable adjectives selected from the materials described in full elsewhere (4, 6). They were used alternately for work and practice. Ten of the words in each list were arbitrarily called wrong and 10 right. Neither the right or wrong words had any distinguishing characteristic that marked them as right or wrong. That is, they were placed in either category by chance. The words were presented to the subject one every 2 sec. by the photographic projection method previously described (6). They appeared in 6 different random orders of presentation so that the sequence of right and wrong words changed from trial to trial. There were never more than 5 right or wrong words in succession.

The subjects were told that some of the words they were going to see were right and some wrong. They were to learn to say the right ones and not to say the wrong ones. They were told they would not be able to tell the right ones from the wrong ones the first time through the list on any *a priori* basis. Rather they were to say some of the words at random. If they happened to say a right word, a bell would ring indicating that they were to continue saying that word whenever they saw it (*i.e.*, that it was right). If they happened to say a wrong word, the bell would not ring and they would know the word was wrong and that they were not to say it from then on. On the other hand if they remained silent in response to certain words, the bell would ring indicating that their response was correct (that the word was wrong and that they were to continue not saying it from then on). Or if they remained silent and the bell did not ring, they would know that their response was incorrect, that the word was right and that they should say it whenever they saw it. In this way they were to build up specific responses of saying certain words in the list and not saying certain others. The bell was rung if the response they made to a particular word was correct. It rang at the end of each 2-second presentation during the fraction of a second when the words were blacked out by a shutter which covered the presentation screen (6).

The subjects were particularly urged not to use any system such as saying all the words and then trying to eliminate those for which the bell did not ring. They were instructed in this way for two reasons—first, to give each word or kind of words approximately equal chance of being responded to correctly on the first presentation, and second, to give approximately an equal amount of practice to both types of response. Equal facility in learning either type of response was especially desired because the experiment was designed to rule out the difference in the degree of learning of the responses to the right and wrong words. Since one list was learned for practice, whenever it was thought during the practice session that the subjects were saying or not saying too many of the words they were stopped and urged to vary their responses to the doubtful words so as to avoid favoring one type of response. The other general instructions were those previously used.

On the experimental day all subjects learned to a criterion of 15 in 20 correct responses. Half of the subjects then continued learning as before up to mastery. The other half continued learning after a rest period of 2 min. which was filled with color-naming at a 2-second rate. The criterion of mastery was one perfect trial. No subject knew he was going to be stopped and asked to name colors or that he would continue learning the list at the end of 2 min. The words were recorded on mimeographed forms in the 6 orders of presentation. The wrong words were starred so that the experimenter knew what response the subject should be making to a given word and consequently whether the bell should be rung or not.

The subjects were 44 Yale undergraduates paid by the hour for their services, each of whom served one practice and one experimental session.

Results

Table 1 presents original learning scores. A comparison of the average trials taken to reach the criterion of 15 in 20 correct responses shows that the two groups of subjects were equated in learning ability up to the time the rest period was introduced. When these scores are compared with similar measures from an earlier experiment using 2 stimulus words, the change in technique appears to have made little difference in the ease of mastering the material. The average trials taken to reach the original learning criterion (10.02) for both groups combined is somewhat higher than the same average (9.50) for the earlier experiment (**6**), but the difference is not significant (critical ratio = .51).

TABLE 1

Mean Original Learning Scores for Two Conditions Subsequently Varying in Amount of Rest Between Learning and Recall

Condition (Rest Interval)	N	Trials to Reach Criterion	Correct Responses on Criterial Trial		Number of Correct Responses per Trial
			Plus Words	Minus Words	
6 sec.	22	10.27	7.82	7.73	11.52
σ_M		1.02	.28	.25	
2 min.	22	9.77	7.82	8.09	11.71
σ_M		1.11	.25	.30	

'Plus words' refers to responses to right words.
'Minus words' refers to responses to wrong words.

Furthermore the degree of mastery of responses to right and wrong words at the time the rest was introduced is substantially the same whether the comparison be made within a condition or between conditions. Since it was so important to know that the degree of learning of the responses to the right and wrong words was equal at the beginning of relearning, the strength of these two kinds of responses in original learning was measured in another way. The average number of correct responses per trial for each condition, which shows how well equated the conditions were in original learning, is given in the last column of Table 1. The average number of correct responses per trial to the plus words alone was 6.05 and 5.58, to the minus words alone was 5.47 and 6.13 for the 6-second and 2-minute rest groups respectively. The differences are not significant or large, indicating a satisfactory equality in degree of learning.

The measures which yielded the main differential reminiscence effect in the transfer experiment are summarized in Table 2 for the

present experiment. They show no such reminiscence effect as was found previously in the responses to the wrong words. The difference in RL_2[2] is especially striking. In the transfer experiment (6, p. 154) the improvement in the minus words in the 2-minute rest group over the control was 3.37 (σdiff = 1.34). In the present instance there is a net loss of −.59 (σdiff = .69) for the same comparison. The difference between comparisons (3.96) is significant (C.R. = 2.62, $P < .01$).[3] On the other hand, the gain after a 2-minute rest for the plus words in the transfer experiment was .12 (σdiff = .77); the loss here is .55 (σdiff = .47) making a difference of .67, the critical ratio of which is only .74. The change in technique affects the wrong

TABLE 2

MEAN RELEARNING SCORES FOR TWO CONDITIONS DIFFERING IN AMOUNT OF REST BETWEEN LEARNING AND RECALL

Condition (Rest Interval)	N	RL_1 Trials to Get 10 in 10 Correct		RL_2 Trials to Reach Original Learning Level	
		Plus Words	Minus Words	Plus Words	Minus Words
6 sec.	22	6.96	8.32	2.00	2.23
σ_M		.96	1.32	.30	.40
2 min.	22	7.32	7.50	2.55	2.82
σ_M		.83	1.15	.36	.56

'Plus words' refers to responses to right words.
'Minus words' refers to responses to wrong words.

words but not the right ones. This is particularly significant in view of the fact that the same words were right and the same words wrong in the two experiments. Whatever produced the reminiscence in the minus words in the transfer experiment is not operating in the present situation.

Table 3 divides the recall scores into two groups, those obtained by subjects who were making a high number of correct responses per trial in original learning (Group A) and those obtained by subjects who were making a low number of correct responses per original learning trial (Group B). Apparently rest has a different effect on these two groups of subjects. It causes reminiscence in Group A and forgetting in Group B.

[2] RL_2 is a score obtained for an individual by counting the number of relearning trials it took him to get at least as many correct responses to either the right or wrong words as he obtained on the criterial trial of original learning.

[3] This expression is to be read: the probability is less than one in 100 that as large a difference *in the obtained direction* could have occurred by chance in samplings of populations of this size. All similar expressions in this paper are to be read the same way unless otherwise noted. C.R. always stands for critical ratio, σ for standard error of the mean.

In Recall 2, rest produced a net gain in Group A of 2.16(C.R. = 2.06, $P < .03$) and a net loss in Group B of .65 (σdiff = 1.20). Therefore the conclusion appears justified that the rest pause will produce reminiscence in Group A and will not in Group B. The further conclusion that the rest actually produces forgetting in Group B is not so certain, the largest critical ratio (for the loss in RC_1) being 1.62, $P < .07$.

One further interesting fact is that the rest clearly differentiated Groups A and B, Group A, when rested, doing significantly better on Recall 1 (15.45 − 12.50 = 2.95) than Group B when rested (C.R. = 2.98, $P < .01$). The two groups were not clearly differentiated on the same trial if no rest was given either group (14.90 − 14.58 = .32).

The different effect of rest on Groups A and B which appears in the recall scores also appears in the relearning scores but not reliably. For example, Group A reached the criterion of mastery on the average 1.89 trials sooner in the rest condition than Group B in the control condition. There appears to be more reminiscence (greater improvement with rest) in Group A than Group B just as in the case of recall.

Granted that reminiscence appears only in Group A the further problem is to determine whether the improvement in this group is localized equally in the responses to plus and minus words. Table 4 provides a provisional answer. In Table 3 the most reliable reminiscence for Group A was in Recall 2. When this effect is analyzed down into its components as in Table 4, it appears that the reminiscence is localized in the responses to the wrong words. The differ-

TABLE 3

MEAN RECALL SCORES FOR TWO TYPES OF LEARNERS UNDER TWO CONDITIONS

Condition (Rest Interval)	N	Mean Correct per OL Trial	RC_1	σM	RC_2	σM
Group A						
6 sec.	10	12.22	14.90	.64	14.30	.84
2 min.	11*	12.34	15.45	.57	16.46	.62
Difference			.55		2.16	
Group B						
6 sec.	12	10.93	14.58	.91	15.25	.87
2 min.	10	10.76	12.50	.81	14.60	.83
Difference			−2.08		−.65	

OL = Original learning; Group A and B refer to those subjects whose average number of correct responses per original learning trial was above or below the mean respectively for the whole group of subjects. RC_1 and RC_2 = Recall 1 and 2, the first 2 trials after the original learning criterion.

* One subject's scores were omitted because his average correct responses per original learning trial was unusually high and made the mean for his condition considerably higher than the comparable mean in the 6 second rest condition.

ence between the number of responses to minus words correctly recalled in the control and rest group is 2.04 and its critical ratio 2.80, $P < .01$. This result gains added significance from the fact that the average number of correct responses to the minus words on the last original learning trial was almost exactly the same in the two conditions. No such reminiscence appears for the plus words on RC_2.

TABLE 4

GROUP A. MEAN NUMBER OF CORRECT RESPONSES ON LAST TRIAL OF ORIGINAL LEARNING AND FIRST TWO TRIALS OF RELEARNING FOR TWO KINDS OF WORDS

	N	OL_2		RC_1		RC_2	
		Plus	Minus	Plus	Minus	Plus	Minus
Control	10	7.70	8.10	7.60	7.30	7.70	6.60
σ_M		.40	.28	.45	.47	.52	.63
2 min. rest	11	8.20	8.00	7.55	7.91	7.82	8.64
σ_M		.28	.40	.36	.44	.48	.36

'Plus' refers to right words.
'Minus' refers to wrong words.
OL_2 = Criterial trial of original learning.
Group A = High average number of correct responses per original learning trial.

It is possible that the difference in the amount of reminiscence for the two kinds of responses is due to an initial superiority of the rest group in learning responses to the minus words. A check which is more adequate than the equation on the last trial of original learning is to compare the average number of correct responses per original learning trial to the minus words for the two conditions. The values are 5.64 and 6.37 for the control and rest conditions respectively, indicating a superiority for the rest group, of which the difference in RC_2 may be just a reflection. To test the dependability of the RC_2 difference the original learning means were equated in two ways. First, by omission of scores the control mean was raised to 5.93 and the rest lowered to 6.03. Under these circumstances the difference in RC_2 dropped to 1.66, C.R. 2.08 ($P < .03$). Second, by omission of the lowest 4 cases in the control group its original learning mean was raised to 6.37, the equivalent of the mean for the rest group. This method of equation seemed more desirable for two reasons. It maximized the condition (high average per trial) which was found to yield reminiscence and it produced the same average per trial as was obtained for the plus words in both conditions (6.34 and 6.33 for the control and rest conditions respectively). With this method of equation the RC_2 difference in favor of the rest group was 1.97, C.R. = 2.24, ($P = .02$). In short, even with the two conditions

carefully equated in original learning, reminiscence is reliably present in the minus word responses and not in the plus word responses.

Differences which support this conclusion were also obtained in RC_1 and for relearning measures but they were not completely reliable. For example, the difference in favor of the *rest* group in trials to get all 10 responses to the minus words correct (RL_1) was 2.23 trials, its critical ratio 1.40, $P < .09$. On the other hand for the plus words responses there was a difference favoring the *control* group of 1.64.[4] In other words the chances are very slight that reminiscence could ever occur in the plus words and almost equally small that reminiscence would *not* occur in the minus words.

As for Group B which showed no reminiscence, a complete breakdown of the data into responses to plus and to minus words revealed no difference in recall or relearning scores for the two conditions.

It was thought possible that those subjects showing a high average per original learning trial *within* one category of responses might show more reminiscence than those with a low average in that category. An analysis of the data for this factor revealed no consistent or reliable difference. That is, those who made a high number of correct responses to plus or to minus words in original learning showed no more reminiscence within the same group of words than those who made a low number of correct responses to the plus or minus words. The explanation of this probably lies in the fact that a high frequency of correct responses within a kind of response may not reflect learning level to any great extent but only an individual bias in favor of saying or not saying most of the words.

Discussion

The major finding of the experiment is that with the new method of presenting the discriminations no reminiscence appears in the group averages either for the right or wrong words. This fact gains added significance in view of the almost exact comparability of the original learning scores by the two methods. It took just as long to master the list to the original learning criterion when one word was removed from each pair as it did when two words were to be discriminated. It also took the same number of trials on the average to relearn the right words by the method of single stimuli to the criterion of one perfect recitation as when the same right words were combined with

[4] The mean trials to get 10 correct responses to the right words for the control group was 5.00; for the rested group 6.64. The chances are less than 3 in 100 that 6.64 is a sample from a population whose true mean is less than the mean for the control group. The method of arriving at this conclusion has been described in full elsewhere (4). The only different assumption made here is that the control group mean is a sample which is no more than $2\sigma_M$ below its true mean. There is therefore considerable statistical justification for the statement that reminiscence will not appear in the right word responses.

new wrong words in the transfer method. The sole difference between the two methods came in the length of time it took to relearn the wrong words. By the transfer method the time was greatly increased. By the method of single stimuli the relearning time was the same as for the right words.

This fact explains why no reminiscence was obtained for the wrong words. According to the performance inhibition theory it was the difficulty of relearning the wrong words, due to the substitution of strange right words, that made the rest effective. In the present experiment there is no difference in relearning difficulty for right and wrong words. Instead, the wrong words are now as easy to relearn as the right words were in the transfer experiment. Consequently, now as then, no reminiscence should be obtained according to a performance inhibition theory.

While the results thus confirm the inhibition theory of reminiscence, they unfortunately do not shed any more light on the validity of the alternative theories. The mistake in the design of the experiment was to make the wrong words as easy to relearn as the right words rather than to make the right words as hard to relearn as the wrong ones. Only in the latter instance could the experiment have tested the other theories.

Despite this drawback the experiment supplied evidence which effectively disposed of one explanation of reminiscence suggested in a previous paper (**6**). It was thought possible that reminiscence might be a characteristic of avoidance responses since it appeared only for the wrong parts of the discrimination. If this was so, it should have appeared again in this experiment in the wrong words. Since it did not, the explanation can no longer be regarded as even a possibility.

There is also accessory evidence for a performance inhibition theory in the experiment. Group A, consisting of subjects who had a high average number of correct responses per original learning trial, showed reminiscence whereas Group B, consisting of subjects with low averages, did not. The logical interpretation of the high average per trial is that these subjects suffered from performance inhibition. That is, they were scoring just under the criterion for some time. This probably means that while the 'true' habit strength was increasing, their performance was not, perhaps due to conflicts in recitation such as have been suggested before (**5**, **6**). The rest served to release them from these conflicts and their performance more nearly approximated true learning. Group B subjects, on the other hand, were not retarded in the same way so that a rest for them made little or no difference. Such an interpretation clearly supports the view that reminiscence is due to release from per-

formance inhibition. Furthermore it is difficult to see how the results could be interpreted to fit any other theory proposed.[5]

The finding that reminiscence depends on a characteristic of the learning (or learner) has implications for other studies of the phenomenon, particularly for serial learning where the same result should be obtained. That is, those subjects who have a high average number of correct responses before they reach a criterion in serial learning should consistently show more reminiscence than those who do not. Though no one has reported using this particular method of finding subjects who show reminiscence, a number of people have pointed out that certain individuals show the effect while others do not under the same conditions (3). The reason may well be that those who show reminiscence are more subject to performance difficulties than are those who show none. Rest would aid the former by allowing performance to approximate learning more nearly and would have no effect on the latter other than the normal one of causing slight forgetting. Averaging the results from these two kinds of subjects might lead to the impression that there is no reminiscence (7). Data from these serial learning studies should be analyzed in the same way as here.

A theory which attributes reminiscence to release from performance inhibition has a number of implications as to the conditions under which the phenomenon will appear. In general it should result in those cases in which the conflict or inhibition is increased in some way in original learning, as it was in this experiment for the subjects in Group A. As a matter of fact, conditions which should increase such conflict have rather uniformly produced reminiscence in other studies, *e.g.*, complicated recitation procedures which require more responses per unit time such as Hovland's (1) and Ward's (8),[6] procedures which require the same responses but in less time (2), partly learned materials which by Hovland's procedure (1) would maximize the conflict between saying errors and anticipating the next word, inexperienced subjects who will have more difficulty with the rapid recitation procedures, etc. In a similar fashion a release from in-

[5] Group A could have been unlucky and Group B lucky in guessing in original learning. That is, performance in discrimination learning is probably the sum of a learning factor and a chance success factor. Suppose Group A had a higher learning factor and a lower chance factor than Group B. This would explain how Group A might have a better score on recall than Group B, supposing the chance factors equalized. But it would not explain why rest made a difference. There is no easily conceivable way in which rest could have equalized the chance factor while no rest did not. On the other hand, if the detrimental factor in Group A is not bad luck but performance inhibition, rest could easily have influenced it in the way required to explain the results.

[6] Melton and Stone (7) did not obtain reminiscence when they adopted a simpler recitation procedure than either Ward or Hovland although all other conditions were intended to be the same as Hovland's (1). (*Cf.* footnote, 7, p. 297.)

hibition theory can be shown to apply to studies with other materials, such as pursuit learning, which have shown reminiscence. In fact the evidence now is so generally in favor of such a theory that the next step would appear to be to attempt to induce reminiscence in large amounts by setting up conditions that will increase performance conflicts in the learning before the rest is introduced.[7]

One final fact from the present experiment needs explanation. Group A which showed reminiscence showed it only in the responses to the wrong words. How can this be explained by the theory proposed? There are two possibilities: (a) either the wrong words, though they were learned as well as the right ones, accumulated more inhibition in reaching the criterion, or (b) they accumulated the same amount of inhibition but more of it was dissipated by rest. Neither alternative supplies a completely satisfactory solution, but the first seems more likely. The reasoning would be that more conflict is involved in learning not to say a word than in learning to say it. There is considerable evidence from the behavior of the subjects that this was the case. Saying the words was the preferred type of response, and the greatest difficulty seemed to be in eliminating the habit of saying a certain word that should not be said. If greater conflict arose in connection with the wrong words, there should be greater difficulty in learning them. Yet on the criterial trial the number of correct responses to right and wrong words is equal. The only explanation of this is that the wrong words are really better learned because more effort must be expended on them, but they do not show it because of the larger inhibition factor. Removal of this inhibition by rest would reveal the real superiority of the wrong words. The only difficulty is that if the wrong words were really harder to learn, this should have shown up as an inferiority somewhere in original learning. But the average score per trial as well as the averages on the criterial trial are the same for right and wrong words. The explanation consequently is not completely satisfactory.

The explanation from the other alternative is hardly more so, however. For example, the color-naming in the rest interval might have served to exhaust the saying response somewhat so that there was an increased tendency to remain silent which would show up as an improvement in the wrong words. The difficulty here is that such an 'exhaustion' should also show up as a decrease in the number of correct responses to the right words, which should be said. But there is no such decrease after rest. So the problem will have to remain unsolved for the present.

[7] It should be noted that the inhibition referred to here is caused by performance difficulties rather than by interfering wrong associations—a fact which differentiates this theory markedly from Hovland's removal of inhibition (1).

Summary

Forty-four subjects were presented with lists of 20 words. They had to learn to respond to half the words (right words) by reading the word aloud and to the other half (wrong words) by remaining silent. After one practice session, they learned a list to a criterion of 15 correct responses out of a possible 20. 22 subjects then continued learning as before until they reached mastery (one perfect trial); the other 22 continued learning after a 2-minute rest period filled with color-naming. From comparisons of the results of the two conditions, the following conclusions appear justified.

(1) The difficulty of acquiring responses to right and wrong words was the same for both groups of subjects in both learning and relearning.

(2) In contradiction to previous results (**6**) no reminiscence appeared in the group averages for the wrong words after a 2-minute rest. This was attributed to the fact that the wrong words were easier to relearn in this experiment, since this was the only major difference caused by the change in technique from the previous experiment. Such a result is entirely consistent with the view that the reminiscence in the previous experiment was caused by release from performance inhibition but does not demonstrate the impossibility of other theories.

(3) When the subjects were divided into two groups depending on whether they made a high or low average number of correct responses per original learning trial, those who made a high average per original learning trial showed reminiscence; those who made a low average did not. This result clearly favored a release from inhibition theory of reminiscence to the extent that the high score per original learning trial indicated a gap between performance and learning which was closed by rest.

(4) The reminiscence obtained for the inhibited learners was localized in the responses to the wrong words. No completely satisfactory explanation of this result was found.

The general conclusion from these and other recent studies is that the evidence points sufficiently definitely towards a release from inhibition theory of reminiscence for work to begin on trying to produce reminiscence experimentally by increasing performance difficulties during the learning before rest is introduced.

(Manuscript received October 7, 1942)

Bibliography

1. Hovland, C. I. Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and distributed practice. *J. exp. Psychol.*, 1938, 22, 201–224.

2. Hovland, C. I. Experimental studies in rote-learning theory. II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **22**, 338–353.
3. McGeoch, G. O. The conditions of reminiscence. *Amer. J. Psychol.*, 1935, **47**, 65–89.
4. McClelland, D. C. Studies in serial verbal discrimination learning. Unpublished Ph.D. thesis, Yale Univ., 1941.
5. McClelland, D. C. Studies in serial verbal discrimination learning. I. Reminiscence with two speeds of pair presentation. *J. exp. Psychol.*, 1942, **31**, 44–56.
6. McClelland, D. C. Studies in serial verbal discrimination learning. II. Retention of responses to right and wrong words in a transfer situation. *J. exp. Psychol.*, 1942, **31**, 149–162.
7. Melton, A. W., & Stone, G. R. The retention of serial lists of adjectives over short time-intervals with varying rates of presentation. *J. exp. Psychol.*, 1942, **30**, 295–310.
8. Ward, L. B. Reminiscence and rote-learning. *Psychol. Monogr.*, 1937, **49**, No. 220.
9. Wheeler, R. H. *The science of psychology.* (2nd ed.) New York: Crowell, 1940.

A STUDY OF BILATERALLY RECORDED ELECTRO-ENCEPHALOGRAMS OF ADULT STUTTERERS [1,2]

BY LEIGH C. DOUGLASS [3]

Denver, Colorado

INTRODUCTION

In the past, observation of peripheral behavior has provided the only means of approach to the explanation of the phenomenon of stuttering. It is only on the basis of these external data that inferences concerning the more central events or states that may possibly be held responsible for the peripheral events have been made.

More recently, direct physiological measures of central (cortical) activity have been made available by the development of electroencephalography (**9**). Exact interpretation of these central phenomena, as observed, is difficult at the present time, and one can only erect probable or possible hypotheses concerning the central 'state' or activity that is represented by any particular segment of electroencephalographic record, as contrasted with that represented by any other segment of such a record.

It has been shown that the electroencephalogram (EEG) varies considerably from person to person (**4**) and within the same person under differing conditions of mental activity (**2**, **19**). Further, it has been demonstrated that particular patterns of the EEG may be presented by persons exhibiting certain disorders of behavior at the moments that the disorders are made manifest (**5**). Because of these facts, as well as because the study of some of the more central neurophysiological processes may throw some light upon fundamental physiological bases of psychological phenomena, electroencephalography has been used as a method of investigation of stuttering.

Five previous investigations have been concerned with the relationship that may or may not exist between certain changes in electrocortical functioning and stuttering. All of these investigations have been related to one theory of stuttering (the 'dominance theory')

[1] From the Laboratory of Physiological Psychology, Department of Psychology, University of Iowa.

[2] The main content of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Department of Psychology, in the Graduate College of the State University of Iowa, August, 1940.

[3] The author wishes to express his appreciation to Dr. John R. Knott, who directed this research.

(18). This theory postulates that there is a lack of dominance of one cerebral hemisphere over the other, so that integration of bilateral muscular action is not possible at moments when such dominance is below a critical level. When such a lack of integration occurs, the continuance of normal speech becomes impossible, and speech behavior known as stuttering results.

Travis and Knott (20), in the first of these investigations, studied only one cerebral hemisphere, the left and presumably dominant one. Nineteen normal speakers and 17 stutterers were recorded during silence and during speech. Leads were taken from the left motor and left visual areas, acting as one bipolar lead which covered more than one-half of the hemisphere. They analyzed their data quantitatively, measuring the duration and the amplitude of each individual wave of potential.

When they expressed their results in terms of the relative number of waves which fell at a given wave duration, they found in the faster bands (*i.e.*, a duration of 0.06 sec. and below, or 16.6 waves per sec. and above) 20 percent of the stutterers' waves and only 7 percent of the normals' waves, during silence. This difference was statistically significant. During speech no change in this band occurred which could further differentiate normal speakers from stutterers.

Studying their data in terms of the mean *duration* and the mean *amplitude* of waves, it was found that comparing stutterers with themselves, the waves were of lesser duration and greater in amplitude in stuttering than in nonstuttering speech, and that they were of greater amplitude in stuttering speech than in silence. Comparing normal speakers with themselves, the waves were of lesser duration in speech than in silence. Comparing stutterers with normals, the waves were of greater duration in nonstuttering speech and of greater amplitude in stuttering speech.

Two other studies of unilaterally recorded EEG's of stutterers have been made, one by Hamstra (7) and the other by Scarbrough (17). Hamstra simultaneously recorded monopolar EEG's from the left occipital and left motor areas, and also a bipolar EEG from the left motor area. His subjects (12 adult stutterers) were recorded with eyes closed, eyes open, and eyes open reading.

When Hamstra analyzed his results in terms of the mean *number of waves per second*, he found that there was a *tendency* for the mean to be greater during stuttering than in nonstuttering speech (but this was not statistically significant), although the variability in mean number of waves per second was significantly greater during stuttering. From these data Hamstra concluded that "no concise characteristic changes of an electrocortical nature can be ascribed as due to that behavior called stuttering" (p. 29) although he indi-

cated that the results from the left monopolar motor lead were suggestive of further experimentation.

Scarbrough attempted to discriminate between stutterers and nonstutterers by means of EEG's secured during silence. Using 20 adult stutterers and 20 adult nonstutterers, all right-handed as to habitual usage, he recorded EEG's simultaneously from monopolar leads over the left frontal, left motor, and left occipital areas. Looking for any statistically significant differences in mean number of waves per second between the stutterers and nonstutterers, he found none. Likewise, there were no significant differences in the mean variabilities of the frequencies for the two groups. Seeking to find whether the differences or lack of differences that might be apparent between his two classes of subjects were the same from time to time, he examined two samples of record for each subject, taken 10 minutes apart, and also a second recording one week later. In neither case were the stutterers more variable than the normals. Three of his stutterers and one of his normals showed wave forms or frequencies that might possibly be called abnormal.

In addition to pointing out the theoretical implications of these negative data, Scarbrough suggested that they had therapeutic significance as well, in that the stutterer should not be treated as though he were neurophysiologically different from the nonstutterer.

The first study of bilaterally recorded EEG's of stutterers was made by Travis and Knott (**21**), who recorded from normal speakers and stutterers in order to determine the degree of synchrony and isomorphism of the two hemispheres appearing in these two groups. As in their previous study, they used widely spaced bipolar leads, motor area to visual area. They defined synchrony in terms of phase relations, requiring that the phase relations of the two EEG's (right and left) should be within 45 degrees. Isomorphism was defined in terms of similarity of pattern of the EEG's of the two hemispheres. They found that stutterers, during nonstuttering speech, showed significantly greater synchrony than normals during their speech. With regard to isomorphism, they found that normals, in silence, showed significantly more isomorphism than stutterers in silence; that stutterers, in nonstuttering speech, showed significantly more isomorphism than did stutterers in silence; and that stutterers, in stuttering speech, showed significantly more isomorphism than did stutterers in silence.

From these data, Travis and Knott hypothesized that one "might assume that since the stutterers were less perfectly matched with regard to similarity (isomorphism) during silence than were the normals, there is some relation between the relative amount of such dissimilarity (lack of isomorphism) in stutterers and the severity of

their stuttering" (p. 241). In order to test this assumption they computed rank order correlations between their measures of isomorphism and synchrony and the percent-time stuttered during the experimental period by each subject as computed from the duration of the moments of stuttering signalled on the record. These correlations indicated that the more severe the stuttering the less synchrony during silence and during stuttering speech and the more severe the stuttering the more isomorphism during silence. Thus, there is an apparent inconsistency in relation to synchrony, since the stutterers were more synchronous during speech than the normals during speech, but the more synchronous a given stutterer's bilateral EEG's the less severe was his stuttering likely to be. This inconsistency is not found in relation to isomorphism of the EEG's of the two hemispheres.

Lindsley (**13**) has presented data on the relationship between stuttering and occipital and motor areas EEG's bilaterally recorded by means of closely spaced bipolar leads. He investigated the factors of synchrony and isomorphism, measuring the absence of the latter in terms of unilateral blocking of the alpha rhythm. Records were taken under the conditions of silence, propositional speech, and speech involving the repetition of short phrases presented by the experimenter. Two subjects were studied, one in four separate experimental sessions and the other in only one. Lindsley found no significant or consistent change in degree of synchrony of the two hemispheres from silence to speech, one subject increasing slightly while the other decreased slightly. However, he found that "usually just prior to each of the marked stuttering periods there is an interval during which the alpha waves of the two hemispheres are completely reversed in phase" (p. 219). Unilateral blocking was increased during speech, especially in the motor areas. Also, during or just prior to some of the observed periods of stuttering there was unilateral blocking of the alpha rhythm. There was a greater amount of unilateral blocking in the left hemisphere in one subject and in the right hemisphere in the other. When Lindsley compared the amounts of unilateral blocking and asynchrony in his two subjects with records obtained relative to laterality and the EEG in children, he found that for each of these measures of the EEG the stutterers more nearly approached the 'norms' for ambidextrous children than those for children with marked laterality. Both of his stutterers were reported to be of mixed laterality.

From these data Lindsley is led to hypothesize that "asynchronism of the alpha waves of bilaterally homologous regions of the cortex implies that impulses cannot arise simultaneously in these regions" (thus making the simultaneous innervation of bilaterally paired speech muscles impossible, so that stuttering results). However,

another interpretation of such out-of-phase relationships can be made, for they can be attributed, until very specific evidence to the contrary is forthcoming, not to any basic excitatory level of cortical action but to the method of recording which was used.

Bipolar electrodes have been used to show phase reversals indicative of a shift in the locale of focus (and to find a focus) of the alpha rhythm (1, 13). Adrian and Yamagiwa used such leads to demonstrate that the focus of the alpha was not fixed but moved over an area of several square centimeters. If the occipital alpha of both hemispheres was being recorded, it would seem possible for the focus in one hemisphere to shift in such a manner that the two EEG's would apparently be out-of-phase; but if more bipolar leads were being taken from each hemisphere, some of them might still be mutually in phase. If all of these pairs showed reversal of phase, then some other cause for the reversal might be more safely assumed.

To summarize the existing data from the studies of the EEG in regard to stutterers and to stuttering: (a) unilateral studies do not show consistently changed EEG's relative to stuttering; (b) bilateral studies are not in agreement, especially in regard to the blocking of the alpha rhythm. Differences in technique, as well as differences in judging what is, and what is not, a moment of stuttering may account for the disparity.

Statement of Problem

The problem to be studied here may be specifically stated as follows: What are the electroencephalographic characteristics of stutterers as opposed to nonstutterers, and of stuttering speech as opposed to nonstuttering speech (of both stutterers and nonstutterers), as those characteristics are measured in terms of the presence or absence of the alpha rhythm? Thus, it is proposed to study the percentage of time an alpha rhythm appears in a given area, the percentage of time that an alpha rhythm appears in one hemisphere but not in the other, and the percentage of time that an alpha rhythm appears in neither one hemisphere nor the other.[4]

Further, what interhemisphere relationships in these characteristics can be determined, as these may be related to speech and stuttering?

Apparatus and Procedure

Electroencephalograms were recorded simultaneously from four cortical areas, the right and left occipital areas (RO and LO) and the right and left motor areas (RM and LM) by means

[4] The primary characteristic of the alpha rhythm not dealt with here is that of interhemisphere phase relationships. It has not been considered at this time for the reason that a separate, detailed examination of the meaning of phase reversal is necessary, as has been pointed out above.

of the monopolar method. The placement of the occipital area leads was at a point 3 cm. above the inion and 3 or 3.5 cm. to the right (or left). The placement of the motor area leads was at a point 3 or 3.5 cm. from the midline, along a line determined by Chiene's method for the localization of the motor area (3). The electrodes were solder disks attached to No. 32 enameled copper wire. The disks were securely attached to the scalp by means of a collodion seal. Commercial electrode jelly was interposed between the disk and the scalp to insure good contact. The 'reference' lead was a pair of electrodes attached in parallel to the lobes of the two ears; all cortical leads were recorded 'common' to these. Four 4-stage resistance-capacitance coupled amplifiers, which were coupled to a Grass 4-element ink-writing oscillograph by means of special matching amplifiers, were used for recording the EEG's. The subjects sat in a comfortable chair in a soundproofed, electrostatically shielded room. This room was within a room adjacent to the recording room. A one-way screen provided constant observation of the subjects by two assistant experimenters. By means of an intercommunication system the subject was in constant verbal contact with all of the experimenters, and the experimenter in charge of the recording apparatus was in selective contact with the subject.

The order of recording was as follows: The subjects, comfortably blindfolded and with eyes closed, were recorded during a period of silence for approximately 5 min. The blindfold was then removed and they were instructed to read, silently, a preselected passage. (No recording was done during this interval.) The blindfold was then replaced, and the eyes were closed; the subjects were then instructed to paraphrase, orally, the material which they had just read. (This made it possible to have all of the subjects talk about the same topic, thus ruling out to a great extent the effects of relatively 'automatic' speech.) During this speech interval recording was continuous.

The two assistant experimenters (who were experienced speech clinicians) observed the subjects during the interval of speech and by means of both visual and auditory cues attempted to judge when stuttering occurred. These judgments were indicated on the record by means of signal circuits which were independently controlled by each judge. In order to obtain similarly sampled intervals of nonstuttering speech, the judges were instructed, by means of a signalling device controlled by the principal experimenter, to change their observations from stuttering to nonstuttering speech.

In the analysis of the total speech record two types of speech were thus indicated. The definition of such types and their occurrence was in terms of the agreement of the judgments made by the two assistant experimenters. In order to obtain sufficiently long intervals of record for quantitative analysis it was arbitrarily decided that these agreements had to be of at least one-second duration.

Subjects

The subjects were divided into two groups. The stuttering group was composed of 20 subjects (18 male, 2 female; mean age 22 years, 8 months) who either were under treatment in the Speech Clinic of the State University of Iowa or had been diagnosed as stutterers by the clinic staff. The nonstuttering (normal) group was composed of 20 adults (18 males, 2 females; mean age 25 years, 4 months) selected from the student body and staff of the University.

Method of Analysis of Data

The records from each cortical area were read in terms of the presence or absence of alpha rhythm, each of which was expressed as a percentage. Alpha rhythm was defined as the occurrence of three or more consecutive waves with an amplitude above 5 microvolts and with frequencies ranging between 8 and 12 per sec. inclusive. The percent time alpha present was computed independently for each of the four areas. The percent time alpha blocked (absent) was computed relative to the contralateral homologous cortical area. For example, when there was an alpha rhythm present in the LO but not in the RO, there was unilateral RO alpha blocking. When an alpha rhythm was present in neither the LO nor the RO, there was bilateral alpha blocking.

These data were statistically evaluated in terms of *intergroup differences* (*i.e.*, stutterers *vs.* nonstutterers, in the analogous conditions) and *intragroup differences* (*i.e.*, stutterers *vs.* stutterers, nonstutterers *vs.* nonstutterers, in the pertinent conditions). Intergroup differences in means

during *silence* were examined by the method of analysis of variance. Intergroup differences in means during *speech* were analyzed by the covariance technique in order to adjust for chance differences in initial means (during silence). Intragroup differences were investigated by analysis of variance applied to those subjects for whom measures were available in *all* conditions (silence, nonstuttering speech, and, in the case of the stutterers, stuttering speech [5]). Interhemisphere relations were examined in terms of the difference between them (*i.e.*, L minus R) in the divisions of the observations; the variance technique was applied to differences.

When three meters of unmarked record were read again for alpha blocking after a period of six weeks, a coefficient of reliability of .95 was obtained when the two readings were correlated

Results

As has just been indicated, the data and their statistical evaluation may be systematically divided into two sections, one treating *intergroup differences*, the other *intragroup differences*. The first of these two evaluations should provide an answer to the question: Are stutterers and nonstutterers significantly different, in terms of the measures applied, during silence and during (stuttering and nonstuttering) speech? The second of these two evaluations should provide an answer to the question: Are normals different during silence as contrasted with speech; and are stutterers different during silence as contrasted with stuttering speech and nonstuttering speech?

Intergroup Differences

The mean percent time alpha present for each of the two groups in each of the four areas (LO, RO, LM, and RM) during silence, during nonstuttering speech (n-st. sp.) and during stuttering speech (st. sp.) is presented in Table 1.

Analysis of variance indicated no significant differences between the measures of percent time alpha present or blocked (bilateral or unilateral) were found during silence.

Comparison of these measures in the two groups by the method of covariance indicates the following measures to have been significantly different. (All other comparisons were not significant at the 5 percent level.)

LO: Normals, in n-st. sp., had a higher percent time alpha present than did stutterers, in n-st. sp. This difference was significant at the 1 percent level.

Normals, in n-st. sp., had a higher percent time alpha present than did stutterers, in st. sp. This difference was significant at the 1 percent level.

[5] Because of artifacts in the records from the motor leads (sweat, muscle action potential, etc.), the normals for whom measures in the motor areas are available are fewer than those for whom measures were available in the occipital area. Double entry of means in the tables presented in Results has thus been made. Similar entry has been made in the case of the stutterers, since some stutterers did not stutter (according to the definition of agreement by two judges), and some did not exhibit nonstuttering speech.

TABLE 1

MEAN PERCENT TIME ALPHA PRESENT

	Stutterers				Normals		
	N	Silence	N-st. Sp.	St. Sp.	N	Silence	N-st. Sp.
LO	17 18	66.81 68.47	54.10	57.04	20	66.91	67.13
RO	17 18	65.65 65.89	57.09	61.11	20	70.52	71.97
LM	17 18	43.55 43.36	32.48	31.94	13	41.71	37.68
RM	17 18	45.50 43.97	32.74	33.44	13	41.16	39.78

RO: Normals, in n-st. sp., had a higher percent time alpha present than did stutterers, in n-st. sp. This difference was significant at the 1 percent level.

Normals, in n-st. sp., had a higher percent time alpha present than did stutterers in st. sp. This difference was significant at the 5 percent level.

LM: Normals, in n-st. sp., had a higher percent time alpha present than stutterers, in n-st. sp. This difference was significant at the 5 percent level.

TABLE 2

VARIANCES OF PERCENT TIME ALPHA PRESENT

	Stutterers				Normals		
	df	Silence	N-st. Sp.	St. Sp.	df	Silence	N-st. Sp.
LO	16 17	495.04 523.26	686.20	645.68	19	476.35	353.00
RO	16 17	479.35 480.68	255.09	427.48	19	389.99	302.22
LM	16 17	399.35 418.01	492.00	364.72	12	444.25	446.96
RM	16 17	416.82 409.98	369.60	318.82	12	311.59	398.82

The variability of percent time alpha present is given in Table 2, which presents the variances for the two groups, in each of the four areas in each of the conditions recorded. Comparisons of these variances indicated that there were no significant differences between normals and stutterers in the pertinent instances.

The mean percent time alpha blocked has been determined in two ways, unilateral and bilateral. *The mean bilateral percent time alpha blocked* (LO & RO and LM & RM) for the two groups is presented in Table 3.

TABLE 3

MEAN BILATERAL PERCENT TIME ALPHA BLOCKED

	Stutterers				Normals		
	N	Silence	N-st. Sp.	St. Sp.	N	Silence	N-st. Sp.
LO & RO	17 18	21.46 20.37	30.87	27.55	20	18.76	17.66
LM & RM	17 18	43.71 43.99	56.78	56.61	13	42.18	50.58

Comparison of these measures in the two groups by the method of covariance indicates the following measures to have been significantly different. (All other comparisons were not significant at the 5 percent level.)

LO & RO: Stutterers, in n-st. sp., had a higher bilateral percent time alpha blocked than did normals, in n-st. sp. This difference was significant at the 1 percent level.

Stutterers, in st. sp., had a higher bilateral percent time alpha blocked than did normals, in n-st. sp. This difference was significant at the 1 percent level.

The variability of bilateral percent time alpha blocked is given in Table 4, which presents the variances for the two groups in each of the two areas to which this measure was applied. It was found that the following measures were significantly different:

LO & RO: Stutterers, in n-st. sp., had a greater variance of this measure than had normals, in n-st. sp. This difference was significant at the 3 percent level.

TABLE 4

VARIANCES OF BILATERAL PERCENT TIME ALPHA BLOCKED

	Stutterers				Normals		
	df	Silence	N-st. Sp.	St. Sp.	df	Silence	N-st. Sp.
LO & RO	16 17	389.78 398.65	560.10	523.59	19	322.15	215.99
LM & RM	16 17	522.16 552.24	628.51	409.73	12	386.19	510.22

TABLE 5

MEAN UNILATERAL PERCENT TIME ALPHA BLOCKED

	Stutterers				Normals		
	N	Silence	N-st. Sp.	St. Sp.	N	Silence	N-st. Sp.
LO	17 18	11.73 11.16	15.03	15.40	20	14.33	15.21
RO	17 18	12.89 13.75	12.04	11.34	20	10.72	10.37
LM	17 18	12.73 12.65	10.74	11.45	13	11.93	11.73
RM	17 18	10.79 12.04	10.48	9.94	13	12.50	9.71

Stutterers, in st. sp., had a greater variance of this measure than had normals, in n-st. sp. This difference was significant at the 1 percent level.

These differences in variance account, at least in part, for the significant differences in the means found above. On the basis of either of these analyses it appears that the respective samples were drawn from different populations.

The mean unilateral percent time alpha blocked, as computed for the four cortical areas studied in the two groups of subjects, is expressed in Table 5.

A comparison of the mean values of the two groups, for each measure, indicated that there were no statistically significant differences between them.

The variability of unilateral percent time alpha blocked is given in Table 6. This table presents the variances for the two groups in

TABLE 6

VARIANCES OF UNILATERAL PERCENT TIME ALPHA BLOCKED

	Stutterers				Normals		
	df	Silence	N-st. Sp.	St. Sp.	df	Silence	N-st. Sp.
LO	16 17	21.63 23.13	44.96	35.55	19	48.32	68.73
RO	16 17	30.30 29.26	68.28	50.89	19	28.98	35.06
LM	16 17	31.16 24.99	30.41	16.29	12	33.88	33.97
RM	16 17	27.06 23.34	89.50	29.77	12	49.91	22.47

each of the measures studied. Only one set of measures proved significantly different for the two groups of subjects.

RM: Stutterers, in n-st. sp., had a greater variance of this measure than had normals, in n-st. sp. This was significant at the 1 percent level.

Interhemisphere relations have been expressed only in terms of percent alpha present, since percent alpha blocked is, in the comparison, its reciprocal.[6] Thus, all statements concerning this relationship hold inversely for unilateral percent time blocked.

The results of these comparisons are presented in Table 7. When

TABLE 7

MEAN DIFFERENCES AND ESTIMATED VARIANCES IN PERCENT TIME ALPHA PRESENT IN THE OCCIPITAL AND MOTOR AREAS OF THE TWO HEMISPHERES (LEFT MINUS RIGHT)

	LO − RO			
	N	Silence	N-st. Sp.	St. Sp.
Stutterers	20 15	1.91 1.89 (47.27)	−2.99 (130.70)	−4.32 (132.52)
Normals	20	−3.92	−4.84	
	LM − RM			
	N	Silence	N-st. Sp.	St. Sp.
Stutterers	15	−1.03 (14.50)	.30 (96.39)	−1.33 (39.76)
Normals	13	−.12 (36.93)	−2.10 (18.64)	

the stutterers were compared with the nonstutterers one set of scores was found to be significant, the LO − RO, in silence. This difference was significant at the 2 percent level by the t test ($t = 2.53$).

Further, a test of the homogeneity of the distributions of stutterers and nonstutterers relative to the positivity or the negativity of their scores was made. Twelve stutterers had positive differences in LO − RO, and 8 had negative scores. It was found that $\chi^2 = 5.01$, indicating that the hypothesis of homogeneity might be rejected at the 3 percent level of confidence.

The comparisons of variance indicated that LM − RM variance was significantly greater in the normals, silence, than in the stutterers, silence (5 percent level); and that the LM − RM variance was significantly greater in the stutterers' nonstuttering speech than in the normals' nonstuttering speech (1 percent level). These sizeable

[6] This inverse relationship results from the fact that bilateral alpha blocking produces an interhemisphere difference of zero; hence, alpha present and unilateral alpha blocked are negatively but perfectly related.

variances in the group of stutterers undoubtedly militated against the establishment of significant differences between the two groups. They appeared, however, to result from deviations from the mean of a limited number of subjects.

Intragroup Differences

Having obtained evidence that, in certain measures of alpha characteristics in certain cortical areas, stutterers and nonstutterers differ significantly, it becomes necessary to determine whether or not the stutterers differ in such measures from condition to condition; and whether or not any parallel changes in the nonstuttering (control) group exist which would tend to vitiate an interpretation of the data based on stutterers alone.

The mean percent time alpha present in normals has already been presented in Table 1; the data subjected to analysis in the present

TABLE 8

PERCENT TIME ALPHA PRESENT: MEANS FOR STUTTERERS WITH COMPLETE MEASUREMENTS

	Stutterers (N = 15)		
	Silence	N-st. Sp.	St. Sp.
LO	65.70	54.97	52.60
RO	63.83	57.96	56.92
LM	41.31	32.97	29.71
RM	42.35	32.68	31.03

manner are presented in Table 8 (see footnote 4). The following differences within the group of stutterers were found to be significant:

LO: In silence, there was a higher percent time alpha present than in n-st. sp. This difference was significant at the 1 percent level.

In silence, there was a higher percent time alpha present than in st. sp. This difference was significant at the 1 percent level.

RO: In silence, there was a higher percent time alpha present than in n-st. sp. This difference was significant at the 5 percent level.

In silence, there was a higher percent time alpha present than in st. sp. This difference was significant at the 1 percent level.

LM: In silence, there was a higher percent time alpha present than in n-st. sp. This difference was significant at the 5 percent level.

In silence, there was a higher percent time alpha present than in st. sp. This difference was significant at the 1 percent level.

RM: In silence, there was a higher percent time alpha present than in n-st. sp. This difference was significant at the 2 percent level.

In silence, there was a higher percent time alpha present than in st. sp. This difference was significant at the 1 percent level.

When similar comparisons were made within the nonstuttering group, no differences at the 5 percent level were found.

TABLE 9

VARIANCES OF PERCENT TIME ALPHA PRESENT FOR STUTTERERS WITH COMPLETE MEASUREMENTS

	Stutterers (N = 15)		
	Silence df = 14	N-st. Sp. df = 14	St. Sp. df = 14
LO	543.39	757.14	641.78
RO	513.82	555.22	496.56
LM	391.37	531.04	380.22
RM	383.07	405.08	298.97

The variability of percent time alpha present for the nonstutterers has been presented in Table 2. Table 9 presents similar data for the 15 stutterers on whom measurements were complete for all three conditions. A comparison of these variances from the intragroup approach showed that there were no differences significant at the 5 percent level in either the stuttering group or the nonstuttering group.

TABLE 10

MEAN BILATERAL PERCENT TIME ALPHA BLOCKED FOR STUTTERERS WITH COMPLETE MEASUREMENTS

	Stutterers (N = 15)		
	Silence	N-st. Sp.	St. Sp.
LO & RO	22.60	30.11	31.66
LM & RM	46.16	56.54	58.98

The mean percent time alpha blocked has, as before, been evaluated from the standpoint of bilateral and unilateral blocking. *The mean bilateral percent time alpha blocked* for the normal group has been presented in Table 3; for the 15 stutterers amenable to this measurement, the means are presented in Table 10. The following differences within the group of stutterers have been found to be significant:

LO & RO: In st. sp., there was a higher bilateral percent time alpha blocked than in silence. This difference was significant at the 1 percent level.

In n-st. sp., there was a higher bilateral percent time alpha blocked than in silence. This difference was significant at the 1 percent level.

It is to be noted that there was not a significant difference between the percent times bilateral blocking in this group of stutterers when stuttering speech was compared with nonstuttering speech.

In the case of the normal group, no significant differences were found.

TABLE 11

VARIANCES OF BILATERAL PERCENT TIME ALPHA BLOCKED FOR STUTTERERS WITH COMPLETE MEASUREMENTS

	Stutterers (N = 15)		
	Silence df = 14	N-st. Sp. df = 14	St. Sp. df = 14
LO & RO	430.75	617.64	520.52
LM & RM	540.14	679.11	428.80

The variability of bilateral percent time alpha blocked for nonstutterers appears in Table 4; for those stutterers who were measured in the three conditions, these are presented in Table 11. Intragroup comparisons yielded no significant differences in variance.

The mean unilateral percent time alpha blocked is summarized, for the normals, in Table 5, and for the stutterers, in Table 12. No significant differences were found when the measures were compared under the various conditions studied.

The variability of unilateral percent time alpha blocked for the 15 stutterers is given in Table 13; for the normals, in Table 6. Com-

TABLE 12

MEAN UNILATERAL PERCENT TIME ALPHA BLOCKED FOR STUTTERERS WITH COMPLETE MEASUREMENTS

	Stutterers (N = 15)		
	Silence	N-st. Sp.	St. Sp.
LO	11.70	14.92	15.74
RO	13.57	11.93	11.41
LM	12.52	10.49	11.31
RM	11.49	10.78	9.99

parisons of the variances within each group indicated that there were no significant differences in the group of normal speakers, but that in the group of stutterers the following were significant:

RM: Stutterers, in n-st. sp., had a greater variance than they had during silence. This difference was significant at the 1 percent level.

Stutterers, in n-st. sp., had a greater variance than during st. sp. This difference was significant at the 5 percent level.

The determination of the significance of differences in silence, stuttering speech and nonstuttering speech of stutterers with relation to the *interhemisphere differences in percent time alpha present*

TABLE 13

VARIANCES OF UNILATERAL PERCENT TIME ALPHA BLOCKED FOR STUTTERERS WITH COMPLETE MEASUREMENTS

	Stutterers (N = 15)		
	Silence df = 14	N-st. Sp. df = 14	St. Sp. df = 14
LO	22.18	51.17	38.60
RO	30.27	74.54	51.25
LM	26.09	33.49	19.09
RM	26.38	97.27	35.37

was made by the analysis of variance, which tested for the homogeneity of the scores. When LO minus RO scores were treated in this way it was found $F = 2.73$; this was not significant at the 5 percent level. It can be stated, therefore, that there was no significant variation in the difference between the two occipital areas in the three conditions studied. In the case of the motor areas, F was less than one by inspection; the difference was not significant.

The normal speakers did not change significantly with respect to this measure when speech was compared with silence. Again F was found to be less than one.

Inspection of the variances, as pointed out in the previous section, indicated that the stutterers during speech tended to deviate from the mean more than did the normals, for whom the comparisons of variances in speech and silence were not significant. When the stutterers' variances of LO − RO in silence were compared with those of nonstuttering and stuttering speech it was found that the differences were both significant at the 5 percent level, the variances in speech being greater than that found for silence. When the stutterers' variances of LM − RM in silence were compared with non-

stuttering and stuttering speech, the difference in the first pair was significant at the 1 percent level, of the second, at the 5 percent level, the variances again being greater in speech than in silence. As mentioned in the intergroup comparisons, these differences are attributable to a small number of cases.

Discussion

The data just reported indicate that there is a higher percent time alpha in both the left and the right occipital areas of nonstutterers during speech than in those areas of stutterers during speech (either stuttering or nonstuttering). The nonstutterers also had a significantly higher percent time alpha present in the left motor area during speech than the stutterers during nonstuttering speech.

The data pertaining to bilateral occipital alpha blocked are inversely related to the above findings, largely because unilateral mean differences were not found; hence, significantly different alpha present scores automatically imposed significantly different bilateral alpha absent scores.

The prominent point seems to be that there was no differentiation between the stuttering speech and the nonstuttering speech of stutterers with respect to these particular measures, which at the present makes it fairly difficult to interpret the stutterer-normal difference in causal terms. It appears, rather, that an end-result has been measured. That is, a decrease in percent time alpha present may possibly be regarded as a symptom of being a stutterer rather than a contributing cause to stuttering.

The only data which have been collected in this investigation which do systematically divide the two groups are those pertaining to the *difference* in the percent time alpha present of the two hemispheres during silence. Stutterers, as a group, tended to have a positive difference when RO percent time alpha was subtracted from LO percent time alpha, while nonstutterers, as a group, tended to have a negative difference. The means of the groups were significantly different, as were the distributions of stutterers and nonstutterers relative to the direction of the differences. The first test of this difference in the groups was made relative to the 15 stutterers on whom measures throughout the entire series of observations were available. Interestingly enough, the addition of the other five raised, rather than lowered, the significance (from the 5 percent level to the 3 percent level on the chi-square test of the homogeneity of the distributions).

These results are especially interesting in the light of the observations of Raney (13), who studied the difference in the EEG's of the left and right hemispheres in a group of identical twins, mirrored with

respect to handedness. He found that the right-handed twins tended to have a greater right motor area percent time alpha present, while the left-handed twins tended to have a greater left motor area percent time alpha. He stressed the point that the difference was emphasized in cases highly selected for 'dominance.'

If two groups of subjects are significantly different on the basis of physiological measures, and if this difference can be related to some other difference (such as laterality, in Raney's case,and being labelled stutterers, in the present case), there must be some physiological meaning to the relationship. The fact that Lindsley (**14**) found a trend for left occipital unilateral blocking to occur in a right-handed group, and right occipital unilateral blocking to occur in a left-handed group of children seems to lend support to some sort of a relationship between bilateral differences in cortical functioning and laterality. What this pattern of neural dynamics *is* presents a problem which has not yet been experimentally determined. Since we know from the experiments of Raney and of Lindsley that peripheral laterality has a central expression in terms of the relative amounts of the alpha rhythm of the left and right cerebral hemispheres, we may however assume that a measure roughly indicative of central laterality has been introduced.

Therefore we may hypothesize, on the basis of the interhemisphere intergroup differences in silence, a difference in the central laterality of the majority of the stutterers and nonstutterers tested. (Whether this difference is inborn and constitutional, or modifiable, the present data cannot indicate.)

Scarbrough's statements concerning the therapeutic significance of his negative data may of necessity be forced to yield to modification, if this formulation is supported by further work.

This formulation also carries with it the implication, relative to the dominance theory of stuttering, that there should be a difference in the representation of this action when stuttering and nonstuttering speech are compared. The definitions of stuttering and non-stuttering used in this study seem to make a test of this implication impossible. Although the definition was operational and depended on the agreement of two observers, it could not, due to its very nature, include all of the temporal sequence of events which are stuttering. The increased interhemisphere similarity of the two groups which was observed during speech might be evaluated at present in terms of the finding of Henry (**8**) that individuals different in (unilateral) EEG characteristics during silence (and psychologically minimal activity) present a significantly more similar set of measures when they are all introduced into a common activity. Thus, relative to the present data, although the two groups of subjects appear to represent hetero-

geneous populations (in terms of their means) during silence, during the common activity of speech they appear to represent a homogeneous population. This does not, of course, mean that specific neurophysiological action relative to specific speech differences may not be expected to appear when finer measures of what is stuttering and what is not stuttering are devised.

Summary and Conclusions

Bilateral occipital and motor area electroencephalograms were recorded from 20 stutterers and 20 nonstutterers during conditions of silence and speech. The subjects were observed during speech, and stuttering and nonstuttering were defined as temporal agreement of two judges as to whether stuttering or nonstuttering occurred. The records of each subject were examined for evidences of bilateral and unilateral blocking of the alpha rhythm. The percent time of such blocking was computed for each subject for each of the three conditions in the experiment.

The data were then analyzed by the methods of analysis of variance and covariance in order to evaluate intergroup and intragroup differences. Variability was evaluated by computing ratios of variance in all pertinent relations.

Intergroup comparisons showed that the stutterers and the normals could not be differentiated on the basis of the mean percent time unilateral blocking of the alpha rhythm, either in silence or in speech. There was, however, a greater bilateral blocking of the rhythm in the occipital areas in stutterers during speech than in the nonstutterers during speech.

Intragroup comparisons indicated no significant differences in the group of normals between silence and speech, but in the stutterers bilateral occipital blocking was significantly greater during speech than during silence. It may be that this intragroup increase in bilateral blocking (as well as the greater amount of bilateral blocking in stutterers than in normals during speech) may be attributed to a greater 'attention' value of speech for stutterers than for normals (**10, 11**). This hypothesis is put forth as being more acceptable than one postulating a basic physiological difference between stutterers and nonstutterers in this respect.

Interhemisphere relations indicate significant differences between stutterers and normals in the unilateral occipital alpha blocking of the two hemispheres during silence. Stutterers tended to show more blocking in the left occipital area during silence and nonstutterers a greater amount of blocking in the right. That this interhemisphere differentiation of the two groups is lost during speech is interpreted

as being in keeping with the findings of Henry that individuals with differing amounts of alpha activity during the resting state become significantly more similar when they engage in psychologically similar activities.

The interhemisphere differences in unilateral blocking during silence which distinguish the group of stutterers from the group of nonstutterers seem to point to the existence of a physiological difference between the two classes of speakers, and in so doing they invite further investigation.

(Manuscript received September 19, 1942)

References

1. Adrian, E. D., & Yamagiwa, K. The origin of the Berger rhythm. *Brain*, 1935, **58**, 323–351.
2. Bagchi, B. K. The adaptation and variability of response of the human brain rhythm. *J. Psychol.*, 1937, **3**, 463–485.
3. *Cunningham's Text-book of anatomy.* (6th ed.) London: Oxford Univ. Press, 1931.
4. Davis, H., & Davis, P. A. Action potentials of the brain in normal persons and in normal states of cerebral activity. *Arch. Neurol. Psychiat., Chicago*, 1936, **36**, 1214–1224.
5. Gibbs, F. A., Gibbs, E. L., & Lennox, W. G. Epilepsy: a paroxysmal cerebral dysrythmia. *Brain*, 1937, **60**, 377–388.
6. Grass, A. M., & Gibbs, F. A. A Fourier transform of the electroencephalogram. *J. Neurophysiol.*, 1938, **1**, 521–526.
7. Hamstra, R. H. Central and peripheral rhythms: their relation to moments of stuttering. Unpublished Master's thesis, Univ. Iowa, 1939.
8. Henry, C. E. Brain potentials during sleep: an investigation of electroencephalographic individual differences and their constancy. Unpublished Doctor's dissertation, Univ. Iowa, 1940.
9. Jasper, H. H. Electrical signs of cortical activity. *Psychol. Bull.*, 1937, **34**, 411–481.
10. Johnson, W., & Knott, J. R. The moment of stuttering. *J. genet. Psychol.*, 1936, **48**, 473–479.
11. Knott, J. R., & Johnson, W. The factor of attention in relation to the moment of stuttering. *J. genet. Psychol.*, 1936, **48**, 479–480.
12. Lindquist, E. F. *Statistical analysis in educational research.* New York: Houghton Mifflin, 1940.
13. Lindsley, D. B. Foci of activity of the alpha rhythm in the human electro-encephalogram. *J. exp. Psychol.*, 1938, **23**, 159–171.
14. Lindsley, D. B. Bilateral differences in brain potentials from the two cerebral hemispheres in relation to laterality and stuttering. *J. exp. Psychol.*, 1940, **26**, 211–225.
15. Raney, E. T. Brain potentials and lateral dominance in identical twins. *J. exp. Psychol.*, 1939, **24**, 21–39.
16. Rubin, M. A. The distribution of the alpha rhythm over the cerebral cortex of normal man. *J. Neurophysiol.*, 1939, **1**, 313–323.
17. Scarbrough, H. E. A quantitative and qualitative analysis of the electroencephalogram of stutterers and nonstutterers. Unpublished Doctor's Dissertation, Univ. Iowa, 1939.
18. Travis, L. E. *Speech Pathology.* New York: Appleton, 1931.
19. Travis, L. E. Brain potentials and the temporal course of consciousness. *J. exp. Psychol.*, 1937, **21**, 302–309.
20. Travis, L. E., & Knott, J. R. Brain potentials from normal speakers and stutterers. *J. Psychol.*, 1936, **2**, 137–150.
21. Travis, L. E., & Knott, J. R. Bilaterally recorded brain potentials from normal speakers and stutterers. *J. Speech Disorders*, 1935, 239–241.
22. Williams, A. C., Jr. Some psychological correlates of the electroencephalogram. *Arch. Psychol.*, N. Y., 1939, No. 240.

MEDIATED GENERALIZATION AND THE INTERPRETATION OF VERBAL BEHAVIOR: III. EXPERIMENTAL STUDY OF ANTONYM GRADIENTS [1]

BY CHARLES N. COFER, MARJORIE G. JANIS, AND MARY M. ROWELL

The George Washington University

In the previous discussion of mediated and non-mediated generalization (1), it was suggested that the former type of generalization might occur along a number of different gradients in addition to those involving synonyms. One of the possible other gradients suggested was that involving antonyms. The present paper reports an experiment designed to determine experimentally whether generalization does occur to antonyms.

In the previous paper, we described the mechanism which, it is assumed, mediates generalization in the case of synonymous words (1, Fig. 1; see also 3, pp. 27–28). In the case of antonyms it is possible that the mediating mechanism is somewhat more complex than is the case with synonyms. To explore this theoretical possibility would require a great deal of discussion, and we merely mention the point here, reserving until a later time the examination of this question.

A previous study by Riess (4) showed that generalization to antonyms does occur. He used the galvanic skin reaction to a bell as the response to be conditioned. The present experiment employed a larger number of antonyms than Riess used, and the method was essentially the same as that used in a previous investigation of homophone and synonym generalization (2).

METHOD

1. *Materials.*—A list of 10 unrelated familiar words (presentation or P-list) was assembled such that from each word an antonym could be derived. Other complicating semantic or homophonic relationships were minimized. According to our hypothesis, the repetition (reinforcement) of these antonyms should increase the reaction potentials of the words in the original (P) list more than would the repetition of an unrelated list (control), and hence recall of the P-list should reflect this difference. Two reinforcement lists were accordingly constructed: (1) the A-R list was composed of 10 words each one of which was an antonym of one word in the P-list; (2) the C-R_1 list was composed of 10 words which were not related to any words in the P-list in any evident or formal way, directly or indirectly. These two lists, along with the

[1] The present paper represents part of a research project being conducted by Dr. J. P. Foley, Jr., and the senior author under a grant-in-aid from the American Association for the Advancement of Science. We are indebted to Dr. Foley for a critical reading of a preliminary copy of this paper.

P-list, may be found in Table 1. In addition a 'Buffer' or practice list (B) and an Equation List (E. L.) were utilized; these two lists were the same as those used previously and may be found in Table 1 of the previous paper (2).

TABLE 1

LISTS OF WORDS USED IN THE EXPERIMENT

The word order (top to bottom) shown is that of the first presentation of each list throughout the experiment. Only the R-lists were changed in order at any time, as described in the previous paper (2). Numbers in parentheses indicate the relationship between words in the A-R list and words in the P-list.

P	A-R	C-R_1
(1) light	(3) early	each
(2) hard	(4) praise	print
(3) late	(7) timid	tepid
(4) blame	(2) soft	sold
(5) dirty	(8) top	tuck
(6) dead	(9) subtract	suppose
(7) brave	(10) deny	deer
(8) bottom	(6) alive	also
(9) add	(5) clean	cleat
(10) affirm	(1) heavy	hedge
Mean no. letters per word 4.7	5.0	4.7
Range in no. letters per word 3–6	3–8	4–7

2. *Procedure* and *Scoring.*—The procedure and scoring used in this experiment were the same as in the previous study (2). Briefly, the procedure was as follows: Every *S* first learned the Buffer List and then the E. L. list. On the basis of E. L. score (used to equate the reinforcement sub-groups for 'learning ability'), the *S* was assigned to a reinforcement procedure. In the A-R sub-group, this consisted of four presentations of the A-R list; in the C-R_1 sub-group, this consisted of four presentations of the C-R_1 list. In both groups, the reinforcement list was presented immediately after *S* wrote his recall of the E. L. list, and the reinforcement list presentations were followed immediately by the presentation of the P-list. The C-R_2 sub-group was an additional control, introduced to determine the effect of the omission of the reinforcement list on the recall of the P-list; it is to be compared with the C-R_1 sub-group. In all respects it was treated the same as the other two sub-groups except that no reinforcement list was presented between the E. L. and the P lists. All other aspects of the procedure and scoring, such as the method of presentation of material, exposure interval, instructions (modified slightly to suit the needs of this study), and general conditions were the same as those previously described (2). That the three sub-groups were adequately equated in terms of E. L. score may be seen from Table 2.

3. *Subjects.*—84 *Ss* were used in this experiment, 54 male and 30 female. They were all enrolled in the first course in psychology and were naïve with respect to the purpose of the experi-

TABLE 2

DATA CONCERNING THE EQUIVALENCE IN 'LEARNING ABILITY' OF THE THREE SUB-GROUPS AS EVIDENCED BY EQUATION LIST SCORE

Sub-group	N	'Equation List' Performance			
		No. Correct	No. Incorrect	Range in no. Correct	Mean no. Correct
A-R	28	179	9	4–9	6.39
C-R_1	28	179	9	4–8	6.39
C-R_2	28	177	0	4–9	6.32

ment. Each sub-group included 28 *Ss*, and the sex ratio, age, and years spent in college were approximately the same for the *Ss* in each sub-group.

Results

The major results of the experiment are presented in Table 3. The only comparison relevant to the problem of mediated generalization is that between the means of the A-R sub-group and of the C-R_1 sub-group. The average recall score for the P-list is 0.61 words greater for the A-R sub-group than for the C-R_1 sub-group. This difference is in accordance with what would be expected if mediated generalization occurs. The significance ratio of this difference is 1.68, indicating that there are 95 chances in 100 that such a difference could not have arisen as a sampling error from a true difference of zero. Although this critical ratio is somewhat below the value ori-

TABLE 3

Analysis of Performance of the 3 Sub-Groups in the Recall of the Presentation (P) List. *N* is 28 for Each Group

Sub-group	No. Correct	No. Incorrect	Range Correct	Mean Correct	σ Mean	Diff. from C-R_1 Mean	σ diff.	Significance ratio	P-value
A-R	168	19	4–9	6.00	.247	.61	.363	1.68	95
C-R_1	151	15	3–8	5.39	.249	—	—	—	—
C-R_2	159	7	4–9	5.68	.236	.29	.355	0.82	79
Diff. between A-R and C-R_2 means: 0.32							.342	0.94	83

narily accepted for statistical certainty, it allows a reasonable degree of confidence that the obtained difference is a real one.[2,3] Thus, Riess' finding (4) of generalization to antonyms would seem to be confirmed by a different method.

The difference between the average number of words recalled from the P-list by the two control sub-groups is 0.29, with the C-R_2 sub-group having the higher mean. This slight difference indicates, if anything, only that the four presentations of the reinforcement list caused a small amount of interference with recall of the P-list in the case of the C-R_1 sub-group. (The medians for these two groups are practically identical.)

[2] At the end of the experiment, the *Ss* were questioned to determine whether they had observed the relationship of antonymity in the words. Three *Ss* gave no evidence of having discerned it, four others apparently saw it in only a few words, and two apparently saw it as an 'after-thought,' *i.e.*, after their recalls had been completed. The other 19 cases apparently knew of the relationship without question. If the 9 *Ss* who did not clearly see the relationship are excluded from the A-R sub-group and 9 *Ss* equivalent to them in E. L. score are also removed from the C-R_1 sub-group, the difference between the two means is increased by a small amount.

[3] The medians for the 3 sub-groups show somewhat smaller differences than the means. The medians for the A-R, C-R_1, and C-R_2 sub-groups, respectively, are: 6.0, 5.71, and 5.70.

We tabulated, for each sub-group, the number of correct responses for each word in the P-list. The resulting distributions showed no consistent differences among the groups.

Incorrect recalls for each sub-group were likewise tabulated and analyzed. Of the 19 incorrect words in the A-R sub-group, 17 were repetitions of words in the reinforcement list. The two other incorrect words were (1) *divide*, which is in the same class (arithmetic operations) with *subtract* (A-R) and *add* (P); (2) *loud*, which is an antonym of *soft*, found in the A-R list. Only one of the 15 incorrect recalls in the C-R_1 sub-group was a word from the reinforcement list (C-R_1). The other incorrect words were antonyms of words in the P-list or possible synonyms or quasi-homophones of words in the reinforcement and P lists. A similar statement may be made about the 7 incorrect recalls of the C-R_2 sub-group, although all of them of necessity were related to words in the P-list.

Summary

An experiment was designed to test the hypothesis that mediated generalization may occur to antonyms of reinforced words. The results tend to suggest that generalization may occur in this way.

(Manuscript received September 24, 1942)

References

1. Cofer, C. N., & Foley, J. P., Jr. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
2. Foley, J. P., Jr., & Cofer, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 168–175.
3. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1938, **45**, 271–299.
4. Riess, B. F. Personal communication.

THE DESIGN AND TESTING OF MULTIPLE AMPLIFIERS FOR ACTION POTENTIAL RECORDING

BY R. C. DAVIS

Indiana University

If two or more amplifiers and recording devices are to be used simultaneously to record electromyograms or electroencephalograms, an obvious and very important requirement is that they be independent of one another in their operation. To secure independence of two amplifiers with different inputs by means of shielding, grounding, and possibly separate voltage supply requires care but presents no new problems. Any reasonable sort of amplifier circuit can be made 'independent' of another under such a condition.

But such a condition of the inputs as that just described is not the one which obtains when two or more amplifiers are connected to the same individual, and a test of independence made with inputs separated will give misleading results and a false sense of security. Since body fluids form an excellent conducting medium, it is clear that electrodes placed anywhere on or in the body have a fairly low impedance path between them, and there needs to be some consideration whether this path between electrodes connected to different amplifiers will not result in destruction of their independence.

'Single sided' amplifiers, with one terminal of each grounded, as used by the writer and others in a number of studies show a satisfactory degree of independence under actual operating conditions. The nature of the input is diagramed in Fig. 1 on the basis of the analysis made in an earlier note (1). So long as the ground is on a truly neutral point each amplifier will respond to the positivity or negativity of the location of its active electrode with respect to the neutral or inactive point, and a wave affecting one amplifier would affect the other only when it is extremely attenuated by dispersal through the body fluid. (The rate of this attenuation is discussed in the earlier note.) Experimental test under operating conditions confirms the independence of such systems.

There are two disadvantages to such a circuit, however. Single sided amplifiers are more affected by outside disturbances than balanced or 'push-pull' amplifiers. Then if it is desired to record from two remote parts of the body simultaneously the electrocardiogram must appear in at least one of the records. The very strong field of the cardiogram, of course, pervades the whole body, and

locations on different limbs (except the two legs) are not isoelectric with respect to it. One pair of electrodes may be made isoelectric with respect to the heart field by placing them both on the same limb. Since the neutral electrode is common to both pairs, this necessarily means that the cardiogram will appear in the other record.

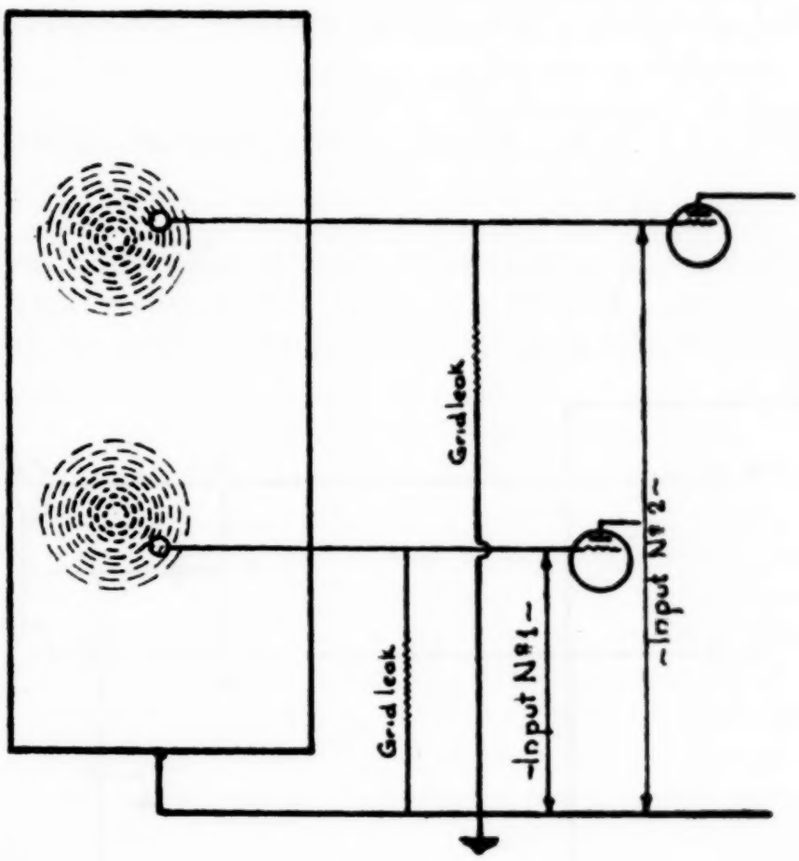
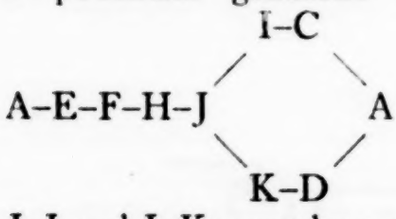


FIG. 1. Input stages of single-sided amplifier connected to the same organism and their relation to a bioelectric potential.

To avoid this often troublesome wave the requirements for an ideal pair of circuits would be that they have no electrode or any other point whatsoever in common. At least one of the recording systems must operate without ground connection to amplifier or oscillograph. Such a circuit was constructed by the writer, using a cathode ray oscillograph tube which required no ground (except for its supply circuits). With a high gain amplifier, stability can be secured only by separating the shielding from the circuit by wide space so as to reduce the capacity. It was found, however, that even after stability had been achieved there was still enough capacity to ground to provide a common point for the two amplifiers and consequently lack of independence. Possibly further reduction of capacity would secure independence.

A circuit frequently used in multiple recording is the standard push-pull design for all amplifiers, with two electrodes for each. In such a case the inputs for two amplifiers are as represented in Fig. 2. (In such an arrangement one electrode of each pair would usually be placed on a region presumed to be neutral.) From this diagram it is clear that a field affecting electrode A would produce a potential gradient (and slight current) through the circuit



In this circuit potential differences across J–I and J–K must be considered for their possible effect on the second recording system. If they are equal and in phase they would, of course, neutralize one another and produce no effect on the amplifier.

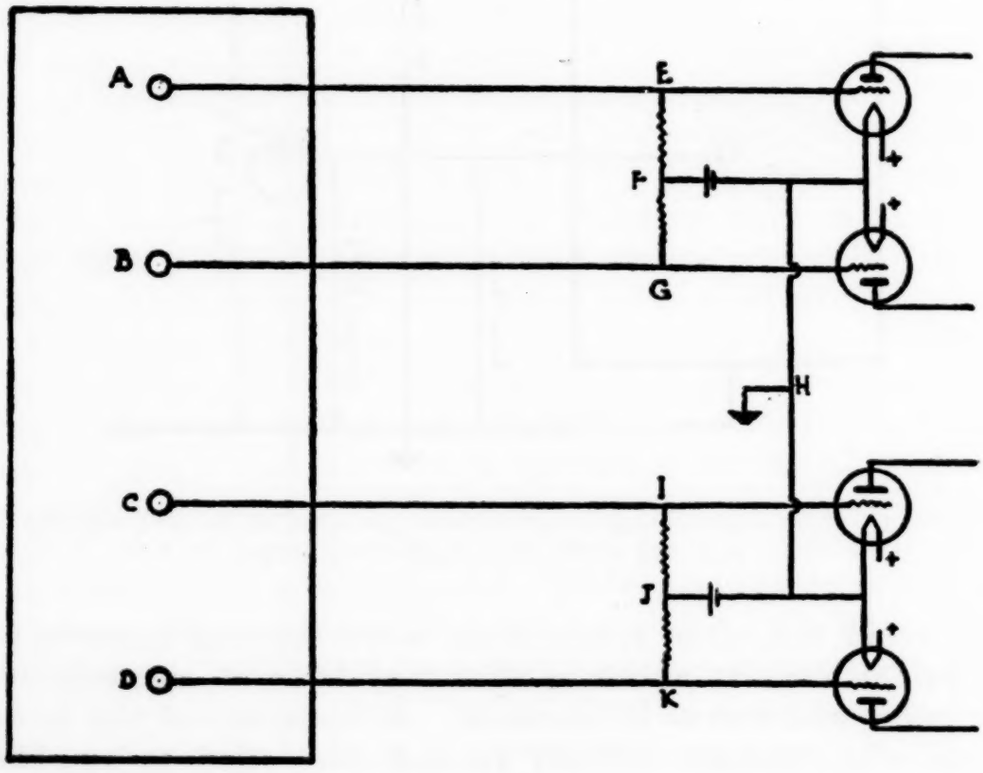


FIG. 2. Input stages of push-pull amplifiers connected to the same organism.

They will be equal and in phase if the two branches of the circuit are equivalent in impedance, *i.e.*, if the path D–A is equivalent to the path C–A. Since this can hardly be assumed for the general case it may be expected that a field arising at any electrode will affect all amplifiers to some degree, the size of the effect being determined by the relationships of the impedances involved in the input circuit.

As a means of breaking the input coupling without removing the

ground from one whole system one may resort to the compromise of isolating one stage of the amplifiers from ground so far as possible, leaving the rest connected in the usual manner. A circuit proposed by Mathews (**2**) consisted of a balanced first stage isolated from ground by a high resistance and feeding into a single sided circuit in succeeding stages. Since this allows part of the input to affect the second stage directly, it produces more 'cross-talk' than a pair of standard 'push-pull' amplifiers (as tested by the writer). An improvement upon this circuit, used in the University of Iowa laboratory, has a regularly grounded push-pull circuit for the succeeding stages (personal communication from John R. Knott).

In tests by the writer it was found that still further isolation could be secured by omitting altogether the high resistance ground connection of the first stage. Stability is apparently just as good without it. In such a circuit, with or without a grounding resistor, there is still a possible coupling circuit through the inputs of the second stages, but its importance should be very much reduced.

The empirical determination of the degree of independence of two systems of a pair must consitute the final test of its merit. For test purposes the electrodes A, B, C, and D were connected by resistors of appropriate value to replace the subject. If, then, a test voltage is introduced across one pair of electrodes, the ratio of the effects it produces on the two amplifiers will constitute a measure of independence under operating conditions. The simplest test potential would be a d.c. pulse taken from a dry cell through an attentuator. To produce a standing wave of a frequency of 100 cps. the writer used the output of an ordinary oscillator fed through an adjustable attenuator and coupled to the input by means of a high grade transformer of 10,000 ohms secondary impedance. It is important, of course, that any test apparatus should not supply a pathway to ground not otherwise present. In the test circuit used, there is, to be sure, the possibility of some back coupling from the secondary to the primary of the transformer and to the grounded oscillator circuit. This was found to be very slight, however, and would in any case tend to make the obtained values somewhat less favorable than they really should be.

With this test circuit it was found that the values of the grid leaks in the first two stages of the amplifiers were important in determining independence. Anything much less than 100,000 ohms in the input would be likely to draw too much current from *S*, and a value of one megohm is so great as to produce interaction. Values of 100,000 ohms were found satisfactory. With the first stages ungrounded (necessitating separate voltage supplies) the ratio between the effects in the two amplifiers was then found to be not greater than .0015 (test voltage = 2.8 volts). For the standard

push pull circuit grounded throughout, the ratio was approximately .0075.

For the most purposes, evidently, either circuit would be satisfactory. When accuracy of 1% or better is required circuits with independent first stages must be employed.

(Manuscript received October 12, 1942)

References

1. Davis, R. C. Properties of electrodes used in recording action potentials from the intact organism. *Amer. J. Psychol.*, 1935, **47**, 693–695.
2. Mathews, B. H. C. A special purpose amplifier. *J. Physiol.*, 1934, **61**, 28P–29P.